ACC. NO. 949

Methuen's Monographs on Physical Subjects

General Editor: B. L. WORSNOP, B.Sc., Ph.D.

WAVE MECHANICS

WAVE MECHANICS

BY

H. T. FLINT, PH.D., D.SC.

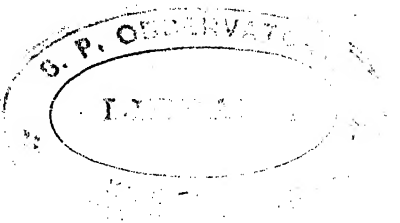
READER IN PHYSICS IN THE UNIVERSITY OF LONDON,
KING'S COLLEGE

WITH A PREFACE BY

O. W. RICHARDSON, F.R.S.

YARROW RESEARCH PROFESSOR OF THE ROYAL SOCIETY
NOBEL LAUREATE IN PHYSICS, 1928

WITH 24 DIAGRAMS



LONDON · METHUEN & CO. LTD.
NEW YORK · JOHN WILEY & SONS

First Published . . . . . *January 31st 1929*
Second Edition, Revised . . *April* *1931*
Third Edition . . . . *September* *1938*
Fourth Edition . . . . *February* *1943*
Fifth Edition . . . . *July* *1946*
Sixth Edition *1949*

CATALOGUE NO. 4035/U (METHUEN)

PRINTED IN GREAT BRITAIN

GENERAL PREFACE

THIS series of small monographs is one which should commend itself to a wide field of readers.

The reader will find in these volumes an up-to-date résumé of the developments in the subjects considered. The references to the standard works and to recent papers will enable him to pursue further those subjects which he finds of especial interest. The monographs should therefore be of great service to physics students who have examinations to consider, to those who are engaged in research in other branches of physics and allied sciences, and to the large number of science masters and others interested in the development of physical science who are no longer in close contact with recent work.

From a consideration of the list of authors it is clear that the reader need have no doubt of the accuracy of the general accounts found in these volumes.

O. W. RICHARDSON

KING'S COLLEGE
June, 1928

PREFACE

THIS little book is written in the hope that it will be of assistance to those who are interested in the most recent development of physics; the new method of approach to atomic problems.

The aim has been to present to the reader in a reasonably simple manner an account of the theory, with its relation to existing physical theory, as developed by de Broglie and Schrödinger. This theory is properly called Wave Mechanics, and it is based on conceptions of continuity. It is in fact a classical theory, for the methods adopted are those which were familiar to physicists before the beginning of this century. The mathematical analysis required has been extensively developed by earlier generations of physicists and may be regarded as part of a familiar equipment. For this reason the theory is approached by ways that the physicist of to-day may not have explored before, but which are well mapped out and amply provided with sign-posts.

No attempt has been made to describe more than one of the two methods of studying these problems. The second has originated with Bohr, Dirac and Heisenberg and is fundamentally different in character, for it is based on a conception of discontinuity and is called The New Quantum Mechanics. The mathematical analysis in this case is less familiar to physicists, is less developed and some of it new.

We do not wish to appear to regard one view as more important or more powerful in dealing with these

problems than the other. It is too early in the history of the new developments to take sides, and to some extent preference for the one or the other view is a matter of taste. It is remarkable that two views so widely different should lead to the same results, and though the reason for this may be easily explained by a comparison of the notations employed, we must not lose sight of the fundamental difference in physical character.

The list of references to original and other works printed at the end of the book is given in the hope that it will be of help to those who wish to consult original papers, and also as an acknowledgment of the sources which have been drawn upon in writing the book. In addition we must acknowledge also two lectures by Prof. E. T. Whittaker given in the course of the last year.

H. T. F.

WHEATSTONE LABORATORY
KING'S COLLEGE
LONDON

December, 1928

PREFACE
TO THE SECOND EDITION

SINCE the first edition of this book was written there have been great advances both in theory and practice in the branch of physics with which it deals.

It has been necessary to make alterations in this edition, to omit certain parts and to add others.

The chapter on the experimental work has been rewritten in order to include some of the newer work and to omit some of the older which was unsatisfactory.

In the theoretical work the principal addition is that of the equations of the first order.

In the past two years a number of important books on this subject have appeared, and there is now a considerable literature at hand for those interested in it.

In a book of this size it is possible to select only a few parts of the subject and to try by means of them to show the character of the new methods applied to the study of fundamental phenomena.

The purpose has been to offer to those who have not yet found time to study the subject in detail, and to those who are beginning it, a short account of the main features, and an introduction to original papers and larger works.

H. T. F.

WHEATSTONE LABORATORY
KING'S COLLEGE
LONDON

March, 1931

CONTENTS

WAVE MECHANICS

INTRODUCTION

UNTIL the beginning of the present century the development of theoretical physics had progressed so far and so successfully that the principles underlying natural phenomena seemed to be understood, and it was expected that the same principles would hold in any new domain of knowledge which might be discovered. Nowadays after nearly thirty years of further research we are familiar with the inadequacy of the wonderful structure which previous generations have built up and which we know as the classical theory. It will be worth while to consider briefly the development of this theory, and to glance at its far-reaching successes before pointing out where it proved inadequate.

It is well known that Newton was the first to work out the laws of dynamics, and this branch of knowledge became a self-contained domain as the result of his work and that of his successors. His law of gravitation immediately opened up a new field for the newly-established science in which it reaped, and still reaps, success after success. In the eighteenth and nineteenth centuries in the hands of mathematicians, astronomers and physicists, mechanics attained such a degree of elegance that the whole science could be founded upon a single principle known as the principle of least action, which was first developed by Maupertuis and later in a quite different way by Hamilton.

The principles of mechanics came to be applied extensively in hydrodynamics, in the theory of sound and

in optics, and in each case with success. Later on the development of statistical mechanics by Maxwell, Boltzmann and Gibbs offered an explanation of the second law of thermodynamics which at first appeared to stand aside from the mechanical theory.

The mechanical theory of optics lost ground and was finally defeated altogether by the wave theory, the reaction occurring about the beginning of the nineteenth century. Later in this century came the electromagnetic theory in which Maxwell combined the results of his predecessors into an exact theory and showed that optics was a branch of the theory of electromagnetism.

The century was distinguished also by the experimental work of Sir J. J. Thomson on the discontinuity of electricity, and by the introduction of the hypothesis of discontinuity into the Maxwellian theory by H. A. Lorentz.

The discovery of positive and negative charges and of their universal occurrence led one school of thought to the electrical theory of matter. If matter is ultimately electricity then the phenomena of Nature may be expected to be ultimately electromagnetic, and the forces experienced are electromagnetic in origin. One might anticipate that mechanics itself would turn out to be a branch of electromagnetism. The introduction of Lorentz's formula for the force on a moving charge to complete Maxwell's equations was a step towards joining up the mechanical and electromagnetic domain.

The state of affairs at the beginning of this century certainly justified the hope that there would be an early blending of these apparently separate domains into a unified physical territory.

The opening years of the century, however, brought to light a very serious difficulty. It proved to be one of many which the classical theory had to encounter. The interest in the union of existing theories was lost in the attempt to account by known principles for the experimental results now clamouring for explanation.

There were originally two clouds darkening the classical sky, with one of which we are not now concerned. This

was the difficulty associated with the Michelson-Morley experiment which has been removed by the principle of relativity. This principle does not actually depart from the methods of classical physics, it embodies the same spirit, and though it has brought an attitude to physical phenomena which is very far removed from that of our classical predecessors, it reveals the older theory as a close approximation to the new. No matter how differently one may regard the law of gravitation in the light of the general principle of relativity we can still describe Newton's law of gravitation as a first approximation and in fact a very close approximation. Nevertheless, it was natural to hope that the modification of Newtonian mechanics which came with the new theory would help to smooth away the difficulties.

The second difficulty indicated a breakdown in statistical mechanics. One of the consequences of this theory is the principle of equipartition of energy. This principle may be understood by considering the case of a molecule which moves without rotation in a collection of similar and similarly moving molecules. The kinetic energy corresponding to the x-co-ordinate of the particle is $\frac{1}{2}m\dot{x}^2$ where $\dot{x}$ denotes the velocity and m the mass of the particle. The principle of equipartition of energy states that the average value of $\frac{1}{2}m\dot{x}^2$ is the same as the average value of $\frac{1}{2}m\dot{y}^2$ and of $\frac{1}{2}m\dot{z}^2$ where y and z denote the other two co-ordinates. This value is in fact $\frac{1}{2}k\mathrm{T}$, where k is Boltzmann's constant and T the absolute temperature.

In this case we have three co-ordinates which fix the position of the particle; in general there may be any number to fix each unit of which the assembly is composed. The principle of equipartition states that the kinetic energy corresponding to each component has on the average the same value $\frac{1}{2}k\mathrm{T}$. This result, which is an unavoidable conclusion from the classical theory, when applied to the problem of determining how the energy in radiation is distributed with respect to the wave-lengths leads to a contradiction with experiment. Although it is in agreement with the experimental findings in the case of

long waves it is inaccurate in the case of short waves and leads to the result that the total energy density of radiation is infinite. This can have no physical meaning.

It was in attempting to explain on theoretical grounds this question concerning black body radiation that Planck introduced his quantum hypothesis which formed the starting-point of the old quantum theory. Another difficulty arose in that the classical theory of specific heats led to the law of Dulong and Petit that the atomic heats of the elements were constant. There were important deviations from this theory which only received an explanation by the use of quantum principles which were additions to the classical theory and which could in no way be deduced from it.

The study of photo-electric phenomena brought to light the difficulty that electrons emitted from metals by radiation incident upon them came away with an energy dependent on the frequency and not upon the intensity of the incident rays. This received a theoretical explanation by Einstein by a principle again foreign to the classical theory.

In this way a theory grew up whose content appeared at variance with accepted views but which was applied to particular problems simultaneously with the classical theory and was therefore looked upon as an additional limitation to that theory.

The state of physics appeared to many who wrote on this subject about that time to be threatened with chaos, since one set of laws appeared to work well in one domain and another quite different set in another domain. H. Poincaré shortly before his death and after the Solvay Congress of 1911 showed that it was necessary to introduce some quite new principle into theoretical physics, viz.: Planck's hypothesis.

The success of this new theory, which we now call the old quantum theory, was not yet so widespread as to make it acceptable to physicists in general, but its most brilliant success awaited it.

In 1913 Bohr published his atomic theory, and so

opened the most striking epoch in the history of spectra. The success of that theory, developed later by a generalisation of his principles by Sommerfeld and W. Wilson, is well known. The application of Bohr's principles has turned the theory of spectra from being a collection of heterogeneous data into a well-ordered branch of physics, and the success has been so extensive and at the same time so accurate in detail that the serious difficulties were often overlooked which have finally caused a revolt, in which Bohr himself has taken a leading part.

As a result of the old quantum theory a corpuscular character was again attributed to radiation, energy being supposed to be associated with it in the form of quanta of magnitude $h\nu$, where ν is the frequency of radiation and h is Planck's constant ($6{\cdot}55 \times 10^{-27}$ units). With these quanta we must associate a mass $\frac{h\nu}{c^2}$ according to the principle that energy has mass, which is a consequence of the theory of relativity.

One of the most striking applications of this is that of A. H. Compton, who considers the problem of the scattering of radiation by electrons as if it were a problem of impact of quanta and electrons. His results are in agreement with experiment. On the other hand, in the phenomena of interference we are compelled to hold to the wave theory of light, and this is true even in the case of Röntgen rays as the crystal phenomena studied by Laue and by Sir William and W. L. Bragg amply show.

These different phenomena show that radiation has apparently both a corpuscular and an extended character, and the time has come to try to unite these apparently opposing aspects.

Two different lines of attack have been made upon these and other difficulties. One of them initiated by Heisenberg sought to describe the phenomena without considering any hypothetical details in the system considered. In observing the behaviour of electrons, protons and atoms we may say that what is observed consists of spectral frequencies, intensities and states of polarisation.

We do not observe positions, velocities or frequencie electronic orbits, and it is in our equations connect these unobserved quantities that difficulties arise. Heis berg proposed to omit these details from his equati and to develop a system of calculation in which o observed quantities occur. This method calls to m the principle of relativity in which there is no mentior the æther. The æther is a medium whose proper escape observation, and in describing phenomena by principle of relativity any mention of the æther is avoid

In classical physics we have an example of a simi method of procedure in thermodynamics, and it is cl from experience that such methods are very power outlasting the detailed methods in which a model assumed.

Heisenberg's method led to the application of t matrix analysis to these problems, and at the same tin guided by the same principles, Dirac applied his notatic

The purpose in this book is to turn our attention the work of de Broglie and Schrödinger. This consi in the association of a wave-like character with wh have hitherto been regarded merely as corpuscles. the old quantum theory of radiation a corpuscular cha acter has been associated with radiation. We ha become familiar with this double aspect in the case radiation. De Broglie's work is an extension of th idea, and as a result of it we have come to associa particles and waves throughout the domain of physics, mechanics as well as in radiation. Inspired by the wo of de Broglie, Schrödinger has established the theo very beautifully from the principles of dynamics. H methods represent a return to classical methods, as w shall see in the following pages.

It will be impossible for us to study in any detail th methods of Heisenberg and Dirac in this small volum Both these writers have introduced a new method c attack upon the fundamental phenomena of physics.

CHAPTER 1

THE VARIATION PRINCIPLES OF DYNAMICS AND OPTICS

WE begin our study by an examination of the variation principle of Maupertuis and Lagrange to which we have referred.

It will be necessary to go over well-known ground in order to find the starting-place for our venture upon the new theory.

It is our purpose to show that dynamics presents a close analogy to geometrical optics, but that while the latter is extended and reinforced by physical optics, the former has received no corresponding extension. This extension is now taking place under the name 'wave mechanics'.

THE PRINCIPLE OF LEAST ACTION

A particle with velocity v (components v_x, v_y, v_z) has mass m, given by $m = \dfrac{m_0}{\left(1 - \dfrac{v^2}{c^2}\right)}$, and momentum $m\mathrm{v}$.

The fundamental equations of Mechanics are:

$$\frac{d}{dt}(mv_x) = \mathrm{X} = -\frac{\partial \mathrm{V}}{\partial x},$$

$$\frac{d}{dt}(mv_y) = \mathrm{Y} = -\frac{\partial \mathrm{V}}{\partial y},$$

$$\frac{d}{dt}(mv_z) = \mathrm{Z} = -\frac{\partial \mathrm{V}}{\partial z},$$

where X, Y and Z are the force components, which are assumed to have a potential V.

Let us consider the variation of the integral:

$$\int_{A_1}^{A_2}(mv_x dx + mv_y dy + mv_z dz)$$

where the path of integration is represented by a path $A_1 A_2$.

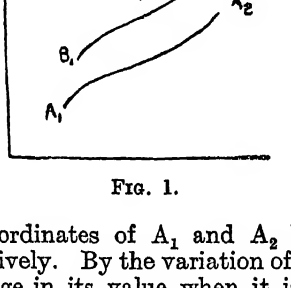


FIG. 1.

Let the co-ordinates of A_1 and A_2 be $(x_1 y_1 z_1)$ and $(x_2 y_2 z_2)$ respectively. By the variation of the integral we mean the change in its value when it is taken over a neighbouring path $B_1 B_2$ (infinitesimally close to $A_1 A_2$) instead of over $A_1 A_2$. This difference or variation is denoted by

$$\delta\int_{A_1}^{A_2} m(v_x dx + v_y dy + v_z dz).$$

Let the co-ordinates of B_1 be $(x_1 + \Delta x_1,\ y_1 + \Delta y_1,\ z_1 + \Delta z_1)$ and of B_2 $(x_2 + \Delta x_2,\ y_2 + \Delta y_2,\ z_2 + \Delta z_2)$.

For convenience we shall write **G** for the momentum $m\mathbf{v}$ and the components are G_x, G_y, G_z.

All the quantities G_x, G_y, G_z and x, y, z are functions of the time, t, and in the operation denoted by δ we suppose that t remains unchanged. This is explained easily by means of a diagram showing two neighbouring curves in a plane (ξ, t).

Let the curve $a_1\ a_2$ represent $\xi = f(t)$, and let $b_1\ b_2$

represent $\xi = f(t) + g(t)$ where $g(t)$ is an infinitesimally small function of t. In considering the value $\delta\xi$ we suppose t to remain constant, i.e. $\delta\xi = g(t) = \mathrm{AB}$ in the diagram.

In the integral dx, dy, dz vary in the passage from one curve to the other, so that in considering the variation of the integral we shall have to take this into account.

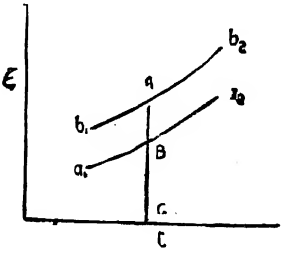


FIG. 2.

For the purpose of evaluating the variation we may consider $\delta\int_{A_1}^{A_2} G_x dx$. If we keep in mind the meaning of the integral, that it is the sum of a series, we see that this expression is the sum of terms of which $\delta(G_x dx)$ is the type. Thus the variation may be written $\Sigma\delta(G_x dx)$, the summation extending over the interval of time which elapses between A_1 and A_2.

Now $\delta(G_x dx) = \delta G_x \,.\, dx + G_x \,.\, \delta dx$.

When a function ξ becomes $\xi + \delta\xi$, $d\xi$ becomes $d(\xi + \delta\xi)$, so that the change in $d\xi$, which by definition is $\delta d\xi$, is equal to $d\delta\xi$.

Hence $\delta(G_x dx) = \delta G_x \,.\, dx + G_x \,.\, d\delta x$.

Thus
$$\delta\int_{A_1}^{A_2} G_x dx = \int_{A_1}^{A_2} \delta G_x dx + \int_{A_1}^{A_2} G_x \,.\, d\delta x$$
$$= \int_{A_1}^{A_2} \delta G_x \,.\, dx + [G_x \,.\, \delta x]_{A_1}^{A_2} - \int_{A_1}^{A_2} dG_x \,.\, \delta x$$
$$= [G_x \delta x]_{A_1}^{A_2} + \int_{A_1}^{A_2} (\delta G_x \,.\, dx - dG_x \,.\, \delta x).$$

For the variation of the integral including the terms in y and z, we have merely to add similar terms to this expression; these we omit for the sake of simplicity.

In the integrated portion of this expression, δx has the values in the limits corresponding to the changes from A_1 to B_1 and from A_2 to B_2 respectively. If the curve A_1A_2 represents a sequence of values which satisfy the mechanical equations, the last expression can be simplified.

In this case $v_x = \frac{dx}{dt}$, $\frac{dG_x}{dt} = -\frac{\partial V}{\partial x}$, etc., the variables x, y, z and t satisfying the curve A_1A_2 as well as these mechanical equations.

Now $$\delta G_x = \delta m \,.\, v_x + m\delta v_x.$$

and since $$m = \frac{m_0}{\left(1 - \frac{v^2}{c^2}\right)^{\frac{1}{2}}}; \quad v^2 = v_x^2 + v_y^2 + v_z^2,$$

where m_0 is the rest mass,

$$\delta m = \frac{m_0}{\left(1 - \frac{v^2}{c^2}\right)^{\frac{3}{2}}}\left(\frac{v_x\delta v_x + v_y\delta v_y + v_z\delta v_z}{c^2}\right),$$

$$\delta G_x \,.\, dx = \delta G_x \,.\, \frac{dx}{dt} \,.\, dt = \delta m \,.\, v_x^2 dt + m v_x \delta v_x \,.\, dt.$$

Thus summing for x, y and z

$$\Sigma\delta G_x \,.\, dx = \delta m \,.\, v^2 dt + m\Sigma v_x \delta v_x \,.\, dt$$

$$= \delta m \,.\, v^2 dt + c^2\left(1 - \frac{v^2}{c^2}\right)\delta m \,.\, dt = c^2 \,.\, \delta m \,.\, dt.$$

We have also

$$\Sigma dG_x \,.\, \delta x = \Sigma \frac{dG_x}{dt} \,.\, \delta x dt = \Sigma\left(-\frac{\partial V}{\partial x}\right)\delta x \,.\, dt.$$

Hence $$\Sigma(\delta G_x \,.\, dx - dG_x \,.\, \delta x) = \left(c^2\delta m + \Sigma\frac{\partial V}{\partial x} \,.\, \delta x\right).\, dt$$

$$= \delta(mc^2 + V) \,.\, dt.$$

Now $mc^2 + V$ is the relativistic expression for the energy of the particle, mc^2 being the energy to be associated with mass m and V the potential energy. We write

$$mc^2 + V = W + m_0c^2,$$

$$\delta\int_{A_1}^{A_2}(G_x dx + G_y dy + G_z dz) =$$

$$\left[G_x\delta x + G_y\delta y + G_z\delta z\right]_{A_1}^{A_2} + \int_{A_1}^{A_2}\delta W \,.\, dt \quad . \quad (1{\cdot}1)$$

This relation holds for the variation of the integral from a mechanically possible path to any neighbouring path.

If now we make A_1 and B_1 coincide and likewise A_2 and B_2, so that the variation is from one path to another possessing the same initial and final points, and if the same energy is associated with both paths so that $\delta W = 0$, then we have

$$\delta\int_{A_1}^{A_2}(G_x dx + G_y dy + G_z dz) = 0.$$

In this case the integral has an extreme value for the mechanically possible path.

If we take the case of Newtonian mechanics where the mass is constant, $G_x dx = mv_x{}^2 dt$, and the integral becomes

$$\delta\int_{1A}^{A_2} 2T dt = 0 \quad . \quad . \quad . \quad (1{\cdot}2)$$

where T is the kinetic energy of the particle.

$\int_{A_1}^{A_2} T dt$ is spoken of as the action of the system and the expression (1·2) is the principle of least action, the oldest variation principle of dynamics. The equation (1·1) is the form in which we must leave the principle in the theory of relativity.

It is worth while to state this principle in words and to point out the very special circumstances under which it applies. The principle of least action states that the integral $\int \Sigma G_x dx$, when taken between two points along

a mechanically possible path, is less than when taken between the same two points along a neighbouring path with which the same amount of energy is associated.

The proof given here shows that the integral has a stationary value, not necessarily a minimum value, and the principle should be described as a stationary principle, not as a minimum principle.

The Hamilton Principle

We may write $G_x dx = G_x \cdot \frac{dx}{dt} \cdot dt$, and consequently

$$\int_{A_1}^{A_2}(G_x dx + G_y dy + G_z dz) = \int_{A_1}^{A_2}(G_x v_x + G_y v_y + G_z v_z)dt$$

$$= \int_{A_1}^{A} \Sigma G_x v_x \cdot dt.$$

Thus from (1·1)

$$\delta\int_{t_1}^{t_2}\Sigma G_x v_x \cdot dt = \Big[\Sigma G_x \delta x\Big]_{t_1}^{t_2} + \int_{t_1}^{t_2}\delta W \cdot dt \quad . \quad (1{\cdot}3)$$

where t_1 and t_2 are the times appropriate to A_1 and A_2 respectively.

But $$\delta\int_{t_1}^{t_2} W dt = W_2\delta t_2 - W_1\delta t_1 + \int_{t_1}^{t_2}\delta W \cdot dt \quad . \quad (1{\cdot}4)$$

where W_1 and W_2 denote the values of W at A_1 and A_2 and δt_1 and δt_2 denote the variations in t from A_1 to B_1 and from A_2 to B_2 respectively. By the principle of energy the value of this quantity is constant for the original path, which is consistent with the mechanical equations, so that $W_1 = W_2 = W$. It should be noted that δW is the change in W in passing from this path to a neighbouring one.

From (1·3) and (1·4) we may deduce directly by writing $L = \Sigma G_x v_x - W$,

$$\delta\int_{t_1}^{t_2} L dt = \Big[\Sigma G_x \delta x\Big]_{t_1}^{t_2} - \Big[W\delta t\Big]_{t_1}^{t_2} \quad . \quad (1{\cdot}5)$$

L is called the Lagrangian function and has a simple interpretation in Newtonian mechanics.

As we have seen, $\Sigma G_x v_x = 2T$ and W is the total energy $T + V$, so that $L = T - V$, the difference between the kinetic and potential energy. In the special relativity mechanics

$$L = \Sigma m v_x^2 - mc^2 - V = m(v^2 - c^2) - V$$
$$= -m_0 c^2 \left(1 - \frac{v^2}{c^2}\right)^{\frac{1}{2}} - V.$$

In the particular case when the variation in (1·5) is referred to two paths with the same initial and final points, i.e. $\delta x_1\ \delta y_1\ \delta z_1$ and $\delta x_2\ \delta y_2\ \delta z_2$ are all zero, the first term on the right of (1·5) vanishes. If in addition the time corresponding to each path is the same so that $\delta t_2 = \delta t_1$, the second term on the right vanishes and we have another integral possessing an extreme value on the mechanical path, viz. $\int_{t_1}^{t_2} L\,dt$.

This theorem, which is known as Hamilton's Principle, is, like the principle of Least Action, subject to close restriction. It is true for the variation from a mechanically possible path to a neighbouring one with the same end points, the time intervals associated with these paths being the same. The addition of a constant to L will not affect the vanishing of the variation under these conditions since $\delta \int_{t_1}^{t_2} (\text{const.})dt = 0$ under the restrictions laid down.

We may thus define L as $m_0 c^2 - m_0 c^2 \left(1 - \frac{v^2}{c^2}\right)^{\frac{1}{2}} - V$ without affecting Hamilton's Principle, and we have the advantage that for small values of v,

$$L = \tfrac{1}{2} m_0 v^2 - V = T - V,$$

or L then becomes the Newtonian expression.

The Variation Principle of Optics

The optical principle corresponding to the mechanics principles just considered is older than these, and is th principle of Fermat. Comparison between them shows very close analogy between these two branches of study a point upon which Hamilton laid stress but which ha been overlooked until recent times. The principle state that a ray of light in passing from a point A_1 to a point A will describe a path for which the time of transit ha a stationary value. If u denote the velocity of light an ds an element of the path,

$$\delta \int_{A_1}^{A_2} \frac{ds}{u} = 0.$$

This relation holds for variation from the optically possible path, just as the corresponding relations i mechanics hold for variation from the mechanically possible path.

The actual value of $\int_{A_1}^{A_2} \frac{ds}{u}$ in this case depends on th co-ordinates of A_1 and A_2, and we may denote it by $V(x_1y_1z_1\ x_2y_2z_2)$.

The value is also determined by the time interval from A_1 to A_2 for

$$\int_{A_1}^{A_2} \frac{ds}{u} = \int_{A_1}^{A_2} dt = t_2 - t_1.$$

We thus have

$$V(x_1y_1z_1\ x_2y_2z_2) = t_2 - t_1 \quad . \quad . \quad (1\cdot6)$$

This is strictly geometrical optics since we are considering a ray, but we can pass from it to the wave theory.

Let A_1 be a fixed point and denote $x_2y_2z_2$ by xyz. Let t_1 be put equal to zero, the times being measured from this instant, and write $t_2 = t$.

The equation (1·6) may now be written

$$V(xyz) = t \quad . \quad . \quad . \quad (1\cdot7)$$

This gives the locus of points which represent the ends of rays starting from a source A_1 which satisfy the optical principle, and may be taken to represent a wave surface at the instant t.

Analogy between the Mechanical and Optical Principles

There must clearly be a wave theory associated with the mechanical theorems corresponding to the wave theory in optics. We have mentioned that this analogy was pointed out by Hamilton, but it has remained to de Broglie and Schrödinger to appreciate the importance of this in the realm of atomic physics. In optics we have extended our theory by means of the wave theory of light, and have built up a satisfactory theoretical basis to account for optical phenomena in the cases where the dimensions are small.

We find no difficulty in accepting as parts of a single consistent theory the subjects of geometrical optics and diffraction. We know exactly what is to be understood by the statement that light travels in straight lines, and we do not feel compelled to speak of classical and non-classical optical theories when we discover a whole range of phenomena in which light bends round edges.

The mechanical wave theory opens up a new field corresponding to diffraction in optics, and we may find a solution of our difficulties in mechanical problems on a small scale by means of our optical analogy. The success already attained in this direction where our practical examples are afforded by the study of the atom has made this branch of mechanics sufficiently important to receive the name of Wave Mechanics.

We may show the derivation of the mechanical wave theory by proceeding from the principle of least action in the same way as we arrive at the optical wave theory from Fermat's principle.

Denote the extreme value of the integral,

$\int_{t_1}^{t_2}(G_x v_x + G_y v_y + G_z v_z)dt$, along the path of which th energy, W, is associated by the function S.

S depends on the co-ordinates of A_1 and A_2 and upon W Thus for the extreme value

$$S(x_1 y_1 z_1, x_2 y_2 z_2, W) = \int_{t_1}^{t_2} \Sigma G_x v_x \, . \, dt$$

and according to (1·1)

$$\delta S = \left[\Sigma G_x \delta x\right]_{A_2} - \left[\Sigma G_x \delta x\right]_{A_1} + \int_{t_1}^{t_2} \delta W \, . \, dt \quad (1\cdot$$

In this process of passage from one path to another w have assumed that the energy, W, associated with th path is constant all along it; for one path the constar value is W, for the other it is W + δW; thus the integr in the last expression is equal to $\delta W(t_2 - t_1)$ since δW a constant change.

Hence

$$\delta S = \left[\Sigma G_x \delta x\right]_{A_2} - \left[\Sigma G_x \delta x\right]_{A} + \delta W(t_2 - t_1) \quad (1\cdot$$

But $$\delta S = \Sigma \frac{\partial S}{\partial x_2} \, . \, \delta x_2 + \Sigma \frac{\partial S}{\partial x_1} \delta x_1 + \frac{\partial S}{\partial W} \, . \, \delta W \quad . \quad (1\cdot 1($$

since by definition S is a function of $x_1 y_1 z_1$, $x_2 y_2 z_2$ and W Thus by comparison of (1·9) and (1·10),

$$\frac{\partial S}{\partial x_2} = G_{x_2} \qquad \frac{\partial S}{\partial x_1} = -G_{x_1} \qquad \frac{\partial S}{\partial W} = t_2 - t_1 \quad (1\cdot 1$$

and there are similar relations in y and z.

Of these we require for our present purpose

$$\frac{\partial S}{\partial W} = t_2 - t_1.$$

As in the optical case, we may consider $x_1y_1z_1$ as our origin, and put $t_1 = 0$ as our time reference. Then $x_2y_2z_2$ is the end of our track and lies on the locus

$$\frac{\partial S}{\partial W} = t \quad . \quad . \quad . \quad (1{\cdot}11)^1$$

corresponding with the optical equation (1·6).

Equation $(1{\cdot}11)^1$ is not derived quite so simply as equation (1·6), nor has it so simple a form, but there is no difference in principle.

We have now a surface associated with our mechanical track, and the mechanical problem is analogous to the optical one.

The same result may be derived from Hamilton's Principle.

In this case we consider the integral $\int_{t_1}^{t_2} L dt$ along a path joining A_1 and A_2 associated with a time interval $(t_2 - t_1)$ which we will denote by t.

Let the extreme value of this integral be denoted by S*, and to show its dependence on $x_1y_1z_1$, $x_2y_2z_2$ and t we write

$$S^*(x_1y_1, z_1, x_2y_2z_2, t) = \int_{t_1}^{t_2} L dt.$$

From (1·5) $\Sigma \frac{\partial S^*}{\partial x_2}\delta x_2 + \Sigma \frac{\partial S^*}{\partial x_1}\delta x_1 + \frac{\partial S^*}{\partial t}\delta t$

$$= \Sigma G_{x_2}\delta x_2 - \Sigma G_{x_2}\delta x_1 - W\delta t$$

Hence

$$\frac{\partial S^*}{\partial x_2} = G_{x_2}, \quad \frac{\partial S^*}{\partial x_1} = -G_{x_2}, \quad \frac{\partial S^*}{\partial t} = -W \qquad (1{\cdot}12)$$

and similar equations in y and z.

Again we fix our attention on one of these, viz.: $\frac{\partial S^*}{\partial t} = -W$, which like $(1{\cdot}11)^1$ gives the locus of $x_2y_2z_2$ associated with a time t measured from A_1 as origin of $(x\,y\,z)$ and $t_1 = 0$ as origin of time.

This extension of mechanics raises a number of interesting questions on the interpretation of material points and quantities associated with them, like energy and momentum. In a wave theory we deal with quantities having extension in space, and while we are familiar with particles, with energy and with momentum in ordinary mechanics we have yet to become familiar with them in the extended sense.

The view taken in wave mechanics is that the material point is a singularity associated with something which is periodic and extended in space; the singularity corresponds to the mass-point of ordinary mechanics.

In the theory of relativity we learn that a quantity of energy, E, is associated with a mass $m = \frac{E}{c^2}$, and it is according to this view unnecessary to distinguish between mass and energy. If E_0 denote the energy of a particle of mass m_0 in a system at rest with respect to the particle, then we know that the energy with respect to a system, with respect to which the particle is moving with velocity v, is given by

$$E = mc^2 = \frac{m_0c^2}{\left(1 - \frac{v^2}{c^2}\right)^{\frac{1}{2}}} = \frac{E_0}{\left(1 - \frac{v^2}{c^2}\right)^{\frac{1}{2}}}.$$

Suppose that the periodicity associated with the particle in the system in which it is at rest is ν_0. It follows from the theory of relativity that in a system, in which it is moving with velocity v, the phenomenon appears as a wave of frequency ν, where ν is given by the formula

$$\nu = \frac{\nu_0}{\left(1 - \frac{v^2}{c^2}\right)^{\frac{1}{2}}}.$$

Thus E and ν change in the same way, and we are led to write

$$E \propto \nu$$
$$\text{or } E = k\nu$$

where k = constant.

Now we know that the equation $E = h\nu$ where $h = 6{\cdot}55 \times 10^{-27}$ C.G.S. units is a relation between energy and frequency which is amply justified in modern experimental physics, so that our constant k may be interpreted as Planck's universal constant h.

There is then no difficulty in understanding how we are to interpret E in the space away from the particle. We must suppose that there is some quantity characteristic of the mechanical wave which at the point which we usually regard as the position of the particle is interpreted physically as the energy of the particle, but which, in the field about this point is interpreted physically as a frequency; the frequency of the mechanical wave.

The momentum is again a quantity for which we have a well-known interpretation when associated with a particle. We have seen that the x-component, G_x, may be made equal to $\frac{\partial S}{\partial x}$. If the surface S is to have a significance in the field corresponding to that of a wave in optics, then the interpretation of momentum in the field becomes clear, it is a vector quantity with components, $\frac{\partial S}{\partial x}, \frac{\partial S}{\partial y}, \frac{\partial S}{\partial z}$, i.e. it is directed along the normal to the surface S = constant.

Another important quantity which we may add to the above is the action. We have seen that in Newtonian mechanics action is measured by $\int \Sigma G_x v_x dt$, and the surfaces, S = constant, are surfaces over which this quantity is a constant. Since we know that action is to be measured in terms of a fundamental unit, h, it is natural to imagine the space filled with surfaces of constant action arranged so that the values of this constant progress from the lowest value h through successive integral values. The constant h provides us with a natural interval for our graduated series of surfaces.

The Canonical Equations and the Hamilton-Jacobi Equation

Let us suppose that the tracks of the particle which we are considering are drawn for short intervals τ and $\tau + \delta\tau$. These times correspond to the track A_1A_2 and the neighbouring varied track B_1B_2 respectively.

We may write for the co-ordinate x_2 of A_2 the following equation:

$$x_2 = x_1 + v_x\tau$$

v_x denoting the x-component of the velocity, supposed uniform, over the small interval τ corresponding to the track A_1A_2.

For the displaced track B_1B_2 we have

$$x_2 + \delta x_2 = x_1 + \delta x_1 + v_x\tau + \delta(v_x\tau)$$

and hence $$\delta x_2 = \delta x_1 + v_x\delta\tau + \tau\delta v_x.$$

Now $S = \int_{t_1}^{t_2} \Sigma G_x v_x \,.\, dt = (\Sigma G_x v_x)\tau$ in this particular case, so that on making the variation

$$\delta S = \Sigma G_x v_x \delta\tau + \Sigma (G_x \delta v_x + \delta G_x \,.\, v_x)\tau \quad . \quad (1{\cdot}13)$$

From equation (1·9)

$$\delta S = \Sigma(G_x\delta x)_2 - \Sigma (G_x\delta x)_1 + \delta W \,.\, \tau$$

and $$(G_x)_2 = (G_x)_1 + (\dot{G}_x)_1\tau,$$

so that the second expression for δS may be written

$$\delta S = (\Sigma\dot{G}_x\delta x - \Sigma G_x\delta v_x - \delta W)\tau + \Sigma G_x v_x \,.\, \delta\tau \quad (1{\cdot}14)$$

and it follows from (1·13) and (1·14) that

$$\delta W = \Sigma v_x \delta G_x - \Sigma\dot{G}_x\delta x \quad . \quad . \quad (1{\cdot}15)$$

From (1·15) we deduce

$$\frac{\partial W}{\partial G_x} = v_x \text{ and } \frac{\partial W}{\partial x} = -\dot{G}_x \,. \quad . \quad (1{\cdot}16)$$

together with two other pairs of equations for y and z.

These six equations are the canonical equations, and

in deriving them we must express W in terms of the co-ordinates and momentum.

It is customary to write $(q_1q_2q_3)$ for the co-ordinates and $(p_1p_2p_3)$ for the corresponding components of momentum, when we may write (1·16) thus :

$$\dot{q}_m = \frac{\partial H}{\partial p_m}, \dot{p}_m = -\frac{\partial H}{\partial q_m}, \quad . \quad . \quad (1\cdot16)^1$$

where m may have the values 1, 2, 3 and H denotes the energy expressed in terms of the co-ordinates and momentum. If, for the sake of simplicity, we limit ourselves to one variable x or q, we may express this by the equation

$$H(q, p) = W.$$

The mechanical problem is thus reduced to the determination of the energy W in terms of the co-ordinates and momentum, the derivation of the canonical equations and their solution.

We may proceed one step farther, but we must content ourselves with a statement of the procedure and omit the argument.

If we denote time by t, we have from the third of the equations (1·12)

$$\frac{\partial S^*}{\partial t} + W = 0$$

and this we may now write

$$\frac{\partial S^*}{\partial t} + H(q, p) = 0.$$

From the first of the equations (1·12) we have $p = \frac{\partial S^*}{\partial x}$ regarding x_2 as any co-ordinate on the track of the particle.

Thus S* must satisfy the equation

$$\frac{\partial S^*}{\partial t} + H\left(q, \frac{\partial S^*}{\partial q}\right) = 0 \quad . \quad . \quad (1\cdot17)$$

In general H depends explicitly on the time, but when this is not the case we may write

$$S^* = -Et + F,$$

where F is some function of the co-ordinates and E is written for a constant appropriate to the problem.

Thus from (1·17) we have

$$H\left(q, \frac{\partial S^*}{\partial q}\right) = H\left(q, \frac{\partial F}{\partial q}\right) = E, \quad . \quad (1·17)^1$$

so that E is the total energy of the system and is constant.

$(1·17)^1$ is the simpler form of the Hamilton-Jacobi equation, (1·17) being the more general form.

The equation is of fundamental importance in classical mechanics and may be used, as we shall see, in the derivation of the wave equation which is the corner-stone of the new theory. The Hamilton-Jacobi equation may be regarded as an approximation to the wave equation.

The importance of this equation is that once a solution of (1·17) is determined in the form $S^*(t, q, \alpha)$, where α is a constant of integration, then the solution of the canonical equations is given by

$$\frac{\partial S^*}{\partial \alpha} = \beta, \quad . \quad . \quad . \quad (1·18)$$

where β is another constant.

In the case we have been considering, viz. that of a particle, we have three co-ordinates and three components of momentum, and there are then three equations like (1·18) containing three constants α and three constants β.

Generalised Co-ordinates

We have used the co-ordinates (x, y, z) in the foregoing pages, and these are the simplest from the physical point of view. They correspond exactly with ordinary physical space. They are, however, not always the most convenient. We shall find them sufficient for our purpose in some cases, but since we have to deal with systems and

not merely mass points we must consider briefly the more convenient methods in use for dealing with the mechanics of systems.

It is found that collections of particles or systems of mass points such as rods, rotators consisting of one particle describing a path round another, and even more complicated structures, can be described with regard to their position and motion by means of a number of independent variables, not of course necessarily x, y and z, the number and nature of these depending on the form of the structure.

The position of a single point is most simply described by stating its distance from three fixed mutually perpendicular planes, by stating in fact its x, y and z co-ordinates. In this case three co-ordinates are necessary and three are sufficient; the particle is said to have three degrees of freedom.

If we wish to fix the position of a rod, AB, we may proceed as follows. Consider a vertical line through its centre of gravity, G, and a plane, S, through this line parallel to some fixed plane. Draw a perpendicular from B on the vertical line and let the foot of the perpendicular be C. If we know the position of the centre of gravity, the inclination of the rod to the vertical and the inclination of the plane GBC to the plane, S, we know the position of the rod exactly, neglecting of course the possibility of rotation of the rod about its long axis.

We thus require five quantities to describe the position of the rod, and these are both necessary and sufficient, any one of them may vary and the others remain unaltered. These five quantities, say, x, y, z, θ and ϕ, are called the generalised co-ordinates of the rod, and we can describe its mechanics by means of them; the rod is also said to have five degrees of freedom.

It is customary to denote the co-ordinates by $q_1 q_2 \ldots q_n$ where n is their number. In the case of a particle each position is given by a set of values (xyz), and in the case of a system each position is given by a set $(q_1 q_2 \ldots q_n)$; the former case is represented by a point in 3-dimensions,

the latter could be represented by analogy by a point in n-dimensions. We should, however, regard the hypothetical n-dimensional space as a mere convenience, we should not give it a physical significance ; the phenomenon is still one of three dimensions.

The kinetic energy of the system can be expressed in terms of the generalised velocities $\frac{dq}{dt}$ or $\dot{q}$, and in the important cases the kinetic energy, T, is a quadratic in the $\dot{q}$'s.

The momentum corresponding to a co-ordinate q_r is defined to be

$$p_r = \frac{\partial \mathrm{T}}{\partial \dot{q}_r},$$

the differentiation being made with respect to $\dot{q}_r$, the other $\dot{q}$'s being regarded as constant in the operation.

These relations for the p's allow us to solve for the $\dot{q}$'s in terms of the p's, and then T can be expressed in terms of the p's instead of the $\dot{q}$'s so that T is a function of the p's and the q's.

This may be illustrated by the case of a particle moving in a plane about the origin, and by describing the position by polar co-ordinates r and θ these will correspond to q_1 and q_2.

The velocity components are $\dot{r}$ and $r\dot{\theta}$, and we have in this case

$$\mathrm{T} = \tfrac{1}{2}m\dot{r}^2 + \tfrac{1}{2}mr^2\dot{\theta}^2,$$

which we may write

$$\mathrm{T} = \tfrac{1}{2}m\dot{q}_1^2 + \tfrac{1}{2}mq_1^2\dot{q}_2^2.$$

By our definition

$$p_1 = \frac{\partial \mathrm{T}}{\partial \dot{q}_1} = m\dot{q}_1$$

$$p_2 = \frac{\partial \mathrm{T}}{\partial \dot{q}_2} = mq_1^2\dot{q}_2.$$

Thus we may write

$$T = \frac{1}{2m} p_1^2 + \frac{1}{2m} \frac{p_2^2}{q_1^2},$$

T in this case being a quadratic function of p_1 and p_2 but containing also q_1. It happens that q_2 (or θ) does not occur in this example, but in general T is a function of all the p's and q's.

There is a very remarkable extension of the theorems we have considered for the case of three co-ordinates to the case of n co-ordinates. The theorem of Maupertuis (1·2) holds for systems as for particles, in fact the case we have considered is a special case of the former.

T must be expressed in terms of the q's and $\dot{q}$'s, and the points A_1 and A_2 which represented two positions $(x_1y_1z_1)$ and $(x_2y_2z_2)$ of the particle now represent two sets of values of the generalised co-ordinates which we may write $(q_1q_2 \ldots q_n)_1$, $(q_1q_2 \ldots q_n)_2$, or we may use our geometrical analogy and say that A_1 and A_2 are two points of our n-dimensional space.

The path of integration was a three-dimensional track along which the particle moved under the mechanical conditions of the problem, and we may now say in a convenient language that the path of integration is along an n-dimensional curve consistent with the mechanics of the system.

In the same way we may determine W, the total energy of the system, and form the expression

$$L = \Sigma p\dot{q} - W,$$

which is analogous to our former expression L in equation (1·5).

Then Hamilton's theorem

$$\delta \int_{t_1}^{t_2} L dt = 0$$

holds for the system as for the particle.

The generalised equations of the system are given by

(1·16)[1], except that m in those equations may have the values 1 to n.

Finally, we have a generalised Hamilton-Jacobi equation (1·17) and a set corresponding to (1·18), in fact if we pass in imagination from 3 to n-dimensions we may pass from the mechanics of a particle to the mechanics of a system.

In making the extension contemplated in wave mechanics we have to take over our geometrical ideas of three dimensions and apply them to n-dimensions. The surfaces corresponding to systems are n-dimensional, where n is equal to the number of degrees of freedom of the system. This we must regard as merely a convenient mode of expression, and when we speak of surfaces and waves in n-dimensions we speak by analogy. The view expressed in the new theory is that waves are to be associated with fundamental units of mass-like protons and electrons, and that these waves have a physical existence. A system is a complicated collection of waves associated with the particles of which the system is composed, but the wave problems concerning systems cannot be solved by considering the individual waves, we have to use what we may describe as the generalised wave theory. It will be seen from this that this generalised wave theory bears to the wave theory of particles the same relation as generalised mechanics bears to the mechanics of a particle.

CHAPTER II

MECHANICAL WAVES

IN this chapter we propose to develop the theory we have just been considering in order to bring out the wave character of mechanics in detail. This is Schrödinger's theory, and we shall finally arrive at the wave equation which represents a new condition to be introduced into physics. We shall develop our theory by considering a property of the surfaces on which S and S* of the last chapter are constant.

By the definition of S and S* it follows that a relation exists between them.

For $S = \int_{t_1}^{t_2} (\Sigma G_x v_x) dt$ and $S^* = \int_{t_1}^{t_2} L dt$

where $L = \Sigma G_x v_x - W.$

Thus $S^* - S = -\int_{t_1}^{t_2} W dt = -W(t_2 - t_1),$

since W is supposed to be constant along the path of integration. If we take the case where $t_1 = 0$ and $t_2 = t$, then

$$S^* = S - Wt \quad . \quad . \quad . \quad (2{\cdot}1)$$

These quantities are very important, and it is worth while to fix our ideas by considering a particular case.

Let a particle of mass m be set in motion vertically from the origin at a time $t = 0$ with velocity u. In this case the origin is the point A_1 and the point A_2 is some

point on the vertical axis. Let the co-ordinate of this point be z, and let the corresponding time be t. In this case

$$S = \int_{A_1}^{A_2} (\Sigma G_x v_x) dt = \int_0^t m\dot{z}\ddot{z}dt = \int_0^z m\dot{z}dz.$$

If g is the acceleration due to gravity $\dot{z}^2 = u^2 - 2gz$, so that

$$S = \int_0^z m(u^2 - 2gz)^{\frac{1}{2}} dz = \frac{mu^3}{3g} - \frac{m}{3g}(u^2 - 2gz)^{\frac{3}{2}}.$$

Since

$$L = T - V = \tfrac{1}{2}m\dot{z}^2 - mgz,$$

$$S^* = \int_0^t \tfrac{1}{2}m(\dot{z} - 2gz)dt.$$

The expression under the sign of integration can readily be expressed in terms of t, and on integrating we find

$$\begin{aligned} S^* &= \frac{mu^3}{3g} - \tfrac{1}{2}mu^2t - \frac{m}{3g}(u - gt)^3 \\ &= \frac{mu^3}{3g} - \frac{m}{3g}(u^2 - 2gz)^{\frac{3}{2}} - \tfrac{1}{2}mu^2t \\ &= S - Wt, \end{aligned}$$

since $W = \frac{1}{2}mu^2$ in this case. This example will serve to show how S and S* can be determined in a special case, and it also illustrates equation (2·1).

Now consider in a general case the surfaces S* = constant. At any time, t, we shall consider a family of such surfaces as filling our three-dimensional space.

In the diagram let us consider three such surfaces, each having the property that S* is constant over it but with a different value of the constant on each.

At any surface S_1* the value of the function S_1* changes with the time, the change in an interval δt being $W\delta t$ by (2·1).

If we consider the surfaces of the diagram as fixed, a particular value S_1* at time t will at time $t + \delta t$ be appropriate to some other surface in the field.

We could draw the surfaces in such a way that a particular value, say S*, would be appropriate to the first at the instant t, to the second at the instant $t + \delta t$, to the third at $t + 2\delta t$, and so on.

It would then appear as if the value S* travelled from one to the other. Since we could make δt as small as we please, we could also describe this by saying that a surface S* travelled outward keeping its value constant as it progressed.

We are familiar with this in the spreading of waves where we have a surface, the wave front, progressing with

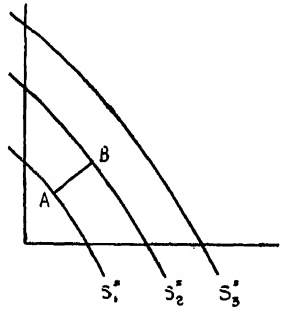


FIG. 3.

the phase constant over it. Let the point A on the diagram have co-ordinates (xyz) and let B be a neighbouring point on the normal to the surface through A with co-ordinates $(x + \delta x, y + \delta y, z + \delta z)$. If S* is the value of the function at A the value at B is

$$S^* + \frac{\partial S^*}{\partial x}\delta x + \frac{\partial S^*}{\partial y}\delta y + \frac{\partial S^*}{\partial z}\delta z \text{ where } \frac{\partial S^*}{\partial x},$$

etc., are components of the rate of change of S* along the normal, these three components combining to give the vector known as the gradient of S* and written: grad S*.

If we write δn for the element along the normal the

change in S* in passing from A to B may be written $\frac{dS^*}{dn}\delta n$, and the length of the vector, grad S*, may be written $\frac{dS^*}{dn}$ or alternatively

$$\left[\left(\frac{\partial S^*}{\partial x}\right)^2 + \left(\frac{\partial S^*}{\partial y}\right)^2 + \left(\frac{\partial S^*}{\partial z}\right)^2\right]^{\frac{1}{2}}.$$

But it will be remembered that $\frac{\partial S^*}{\partial x} = G_x$, the component of momentum, and we have, therefore, the relation

$$\frac{dS^*}{dn} = G \quad . \quad . \quad . \quad (2{\cdot}2)$$

where G is the magnitude of the total momentum, i.e. $(G_x^2 + G_y^2 + G_z^2)^{\frac{1}{2}}$.

We may write

$$\delta n = \frac{\delta S^*}{G}.$$

Now suppose that at A at time t the value for the surface is S*, then in order that it may appear that S* arrives at B with the same surface value at time $t + \delta t$ the condition is that the value appropriate to B is S* at this time. Thus the value at B at time t must be $S^* + W\delta t$ for then at time $t + \delta t$ it becomes $S^* - W\delta t + W\delta t$ or S*. (2·1).

Now we have seen that at a fixed time t, when the value at A is S*, it is $S^* + \frac{dS^*}{dn}.\delta n$ at B, and so our condition is that

$$S^* + \frac{dS^*}{dn}.\delta n = S^* + W\delta t$$

i.e.

$$\frac{\delta n}{\delta t} = \frac{W}{G}.$$

Now $\frac{\delta n}{\delta t}$ represents the velocity with which the waves travel outward along the normal to their front. This is

the velocity which in optics is known as the phase velocity, and since we have seen that the difference in the mechanical wave is merely that S* remains constant instead of the phase, we shall describe this velocity as the phase velocity of the mechanical waves and write

$$U = \frac{W}{G} \quad . \quad . \quad . \quad (2{\cdot}3)$$

We have now to take into account the possibility that these mechanical waves have an actual existence, and if we admit this we have a field open to us in a new branch of mechanics which we know how to approach from our corresponding experience in optics.

Consideration of a Simple Form of Mechanical Wave

The most familiar form of wave in physics is

$$y = a \sin 2\pi\left(\nu t - \frac{x}{\lambda} + \alpha\right) \quad . \quad . \quad (2{\cdot}4)$$

where α is a constant, ν the frequency, λ the wave-length and $2\pi\alpha$ the epoch, i.e. the phase at the instant $t = 0$, at the point $x = 0$. The expression $2\pi\left(\nu t - \frac{x}{\lambda} + \alpha\right)$ is called the phase.

There is a similarity between this and the expression for S* which is the phase in our mechanical waves, for it consists of a function of the co-ordinates, $-\frac{2\pi}{\lambda}x$, and a term containing t, $2\pi\nu t$. These correspond to S and $-Wt$ in the expression for S*, viz.

$$S^* = S - W(t - t_0)$$

while α corresponds to Wt_0.

The simplest form of wave we can obtain by analogy is thus

$$y = A \sin kS^*,$$

where k is a constant and has the dimensions which make $k\mathrm{S}^*$ a quantity without dimensions like the phase in the above wave.

S^* is of the dimensions of action, and we may thus write $k = \frac{2\pi}{h}$ where h is a constant of the dimensions of action.

Thus the simple sine form of wave may be written

$$y = \mathrm{A} \sin \frac{2\pi}{h}(\mathrm{S} - \mathrm{W}t + \mathrm{W}t_0) \quad . \quad . \quad (2{\cdot}5)$$

By its definition S is a function of the co-ordinates and W.

By comparison with (2·4) it is evident that the frequency of this wave is

$$\nu = \frac{\mathrm{W}}{h} \quad . \quad . \quad . \quad . \quad (2{\cdot}6)$$

Thus the mechanical quantity, W, which we interpret as the energy in dynamics, is to be interpreted as $h\nu$ in wave mechanics.

The equation (2·6) has a well-known physical significance in the quantum theory of radiation.

We have seen that the wave theory requires us to recognise that the quantities like action, energy and momentum have a significance at points away from the particle. Equation (2·6) shows that the interpretation of energy in the field is to be $h \times$ frequency of mechanical wave.

Wave Groups

Let two simple waves of the same amplitude, a, but with slightly differing frequencies ν and ν' $(= \nu + d\nu)$ and velocities U and U′ $(= \mathrm{U} + d\mathrm{U})$ travel simultaneously through a medium. They superimpose and produce a disturbance given by

$$\begin{aligned} y &= a\left[\sin 2\pi\left(\nu' t - \frac{\nu'}{\mathrm{U}'}x\right) + \sin 2\pi\left(\nu t - \frac{\nu}{\mathrm{U}}x\right)\right] \\ &= 2a \sin 2\pi\left[\frac{\nu' + \nu}{2}t - \frac{\nu'}{2}\left(\frac{1}{\mathrm{U}'} + \frac{1}{\mathrm{U}}\right)x\right] \\ &\qquad \cos 2\pi\left[\frac{\nu' - \nu}{2}t - \frac{x}{2}\left(\frac{\nu'}{\mathrm{U}'} - \frac{\nu}{\mathrm{U}}\right)\right]. \end{aligned}$$

But $\nu' + \nu = 2\nu + d\nu$ and, neglecting $d\nu$ in comparison with ν, we write $\nu' + \nu = 2\nu$. In the same way $\nu'\left(\frac{1}{U'} + \frac{1}{U}\right)$ may be written $2\frac{\nu}{U}$. Under the cosine the quantities are of the order $d\nu$ so that first differentials must be retained.

Thus $y = 2a \sin 2\pi\left(\nu t - \frac{\nu}{U}x\right) \cos \pi\left[td\nu - xd\left(\frac{\nu}{U}\right)x\right]$.

We may regard y as a wave of frequency ν and of amplitude $a \cos \pi\left\{td\nu - xd\left(\frac{\nu}{U}\right)\right\}$.

The amplitude varies very slowly in comparison with the frequency, ν, of the oscillations. The dotted curve of the figure represents the cosine term, while the full line represents the wave vibrations.

FIG. 4.

If we have a collection of waves with frequencies lying over the small interval ν to $\nu + d\nu$, we represent the vibration y by $ad\nu \sin 2\pi\left(\nu t - \frac{\nu}{U}x\right)$, where a is the average amplitude for this range of frequencies.

For any finite group of waves we have

$$y = \int_{\nu_1}^{\nu_2} a \sin 2\pi\left(\nu t - \frac{\nu}{U}x\right)d\nu, \quad . \quad . \quad (2{\cdot}7)$$

and a will in general depend on the frequency.

The dotted curve in the diagram travels with a velocity given by

$$v = \frac{d\nu}{d\left(\frac{\nu}{U}\right)} = U - \lambda\frac{dU}{d\lambda} \quad . \quad . \quad (2{\cdot}8)$$

This can be understood by examining the cosine term which represents a wave travelling with the velocity v.

λ is the wave-length corresponding to the frequency ν and v is called the group velocity.

We may apply a similar discussion to the mechanical waves. In this case the range $d\nu$ is equal to $\frac{dW}{h}$ by (2·6).

For any two points A_1 and A_2 there is a series of mechanical paths corresponding to the series of values of W lying in the range W to $W + dW$. A group of waves corresponds to this series which, by comparing with (2·7), may be seen to be represented by

$$y = \int_{W_1}^{W_2} a \sin \frac{2\pi S^*}{h} dW \quad . \quad (2\cdot9)$$

where a, which in the optical case depends on the frequency, now depends on W and through this on the frequency of the mechanical wave. The group velocity in this case is

$$v = \frac{d\nu}{d\left(\frac{\nu}{U}\right)} = \frac{dW}{d\left(\frac{W}{U}\right)} = \frac{dW}{dG} \text{ (by 2·3)} \quad . \quad (2\cdot10)$$

Now we have seen that $W = mc^2 + V$, $m = \frac{m_0}{\left(1 - \frac{v^2}{c^2}\right)^{\frac{1}{2}}}$

and it follows that $W = m_0c^2\left(1 + \frac{G^2}{m_0^2c^2}\right)^{\frac{1}{2}} + V$, hence remembering that the change in W which we contemplate is to be made from the first part to the adjacent second path without changing the co-ordinates, of which V is a function, we find

$$\frac{dW}{dG} = v$$

where v is the particle velocity. Thus the group velocity is the same as the velocity of the particle, which we anticipated in (2·10) by denoting the group velocity by v.

The phase velocity, $\frac{W}{G}$, is of course different from this, and we may consider the relation between them most simply by taking the case where there is no field of force.

In this case $W = mc^2$ and $G = mv$, so that

$$U = \frac{W}{G} = \frac{c^2}{v}$$

or

$$Uv = c^2 \quad . \quad . \quad . \quad . \quad (2{\cdot}11)$$

Since v cannot be greater than c, U cannot be less. We therefore know that energy cannot be transported by the phase waves, since according to the principle of relativity energy cannot be transported with a velocity greater than that of light. The phase wave is, however, no more than a mathematical conception, the physicist contemplates groups of waves whenever energy is transported by wave motion so that energy travels with the group velocity. The particle is thus to be regarded as a group of waves and the particle velocity is, in the wave theory, to be interpreted as the velocity of this group.

Diffraction of the Mechanical Waves

We pass on to consider the group of waves (2·9) writing

$$\phi = S^* = S - Wt + Wt_0$$

and making the assumption that the interval over which W extends, i.e. the interval W_1 to W_2, is small.

For any particular mechanical track, which we may by analogy describe as a mechanical ray, there is a particular value of W, and consequently a particular value of ϕ for the mechanical wave.

Thus as we pass from wave to wave of the group, ϕ will vary with W, and consequently

$$d\phi = \frac{\partial \phi}{\partial W} dW \quad . \quad . \quad . \quad (2{\cdot}12)$$

Thus expressing the integral in terms of ϕ,

$$y = \int_{\phi_1}^{\phi_2} \frac{a \sin \frac{2\pi}{h}\phi}{\frac{\partial\phi}{\partial W}} d\phi,$$

where a also depends on W, just as in the optical case it depends on the frequency.

If the range of integration is small we can write

$$y = \frac{\bar{a}}{\frac{\partial\phi}{\partial W}} \int_{\phi_1}^{\phi_2} \sin \frac{2\pi}{h}\phi d\phi,$$

where the bars denote average values over the range.

The value of the integral is

$$\frac{h}{2\pi}\left(\cos \frac{2\pi}{h}\phi_1 - \cos \frac{2\pi}{h}\phi_2\right) = \frac{h}{\pi} \sin \frac{\pi}{h}(\phi_2 - \phi_1) \sin \frac{\pi}{h}(\phi_2 + \phi_1).$$

We take the range to be small enough to write $\frac{\phi_2 + \phi_1}{2} = \phi$ (an average for the extent), while $\phi_2 - \phi_1 = \Delta\phi$, the range.

Thus
$$y = \frac{\bar{a}}{\frac{\partial\phi}{\partial W}} \frac{\sin\left(\frac{\pi}{h}\Delta\phi\right)}{\frac{\pi}{h}} \sin \frac{2\pi}{h}\phi \quad . \quad (2{\cdot}13)$$

$$= A \sin \frac{2\pi\phi}{h},$$

where A is written for the coefficient of $\sin \frac{2\pi}{h}\phi$.

The group of waves is thus compounded into a wave of phase $\frac{2\pi\phi}{h}$ with a varying amplitude. We have a carrier wave which is modulated.

Expressing $\Delta\phi$ in terms of ΔW by (2·12) we find

$$A = \bar{a}\frac{\sin\left(\frac{\pi}{\bar{h}}\overline{\frac{\partial\phi}{\partial W}}\Delta W\right)}{\frac{\pi}{\bar{h}}\overline{\frac{\partial\phi}{\partial W}}}.$$

If $\overline{\frac{\partial\phi}{\partial W}} = 0$, A has a maximum value $\bar{a}\Delta W$.

This may be verified by writing $\frac{\pi}{\bar{h}}\overline{\frac{\partial\phi}{\partial W}} = x$ and considering the maximum and minimum values of

$$A = \bar{a}\frac{\sin \alpha x}{x} \quad . \quad . \quad . \quad (2\cdot14)$$

where α is written for ΔW.

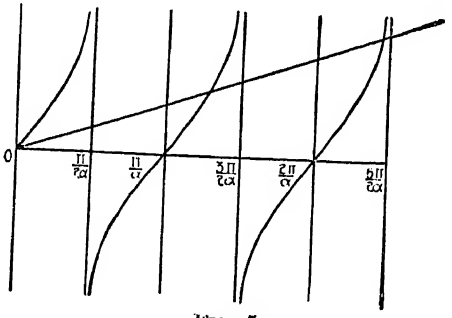


FIG. 5.

This expression occurs in the theory of diffraction in the calculation of the effect of superposing a series of simple harmonic vibrations of equal amplitudes and periods with phases increasing in arithmetical progression. The integration method employed here is shorter.

It follows from (2·14) that maxima occur at points for which the value of x is given by

$$\tan \alpha x = \alpha x.$$

The solution of this may be obtained graphically by noting the intersections of the curve $y = \tan \alpha x$, and the line $y = \alpha x$.

The actual values of x for these intersections are not necessary for our purpose, but they are given in works on Optics.[1] From the graph it is clear that the values of αx tend to $(n + \frac{1}{2})\pi$ as n becomes large.

If the values of A are plotted, the graph obtained is as illustrated in the diagram.

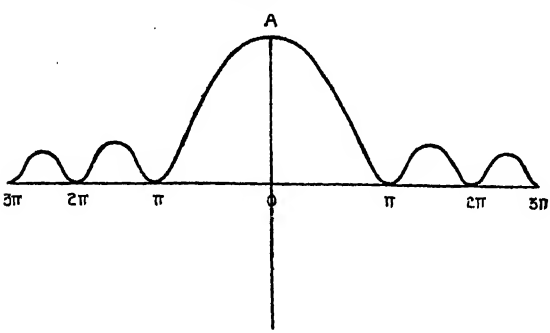


FIG. 6.

We have, therefore, in addition to the maximum $\bar{a}\Delta W$, corresponding to $\frac{\overline{\partial\phi}}{\partial W} = 0$, a series of diminishing maxima at points corresponding approximately to $\frac{\overline{\partial\phi}}{\partial W}\Delta W = (n + \frac{1}{2})h$ with a series of zero values at points where $\frac{\overline{\partial\phi}}{\partial W}\Delta W = nh$, where $n = 1, 2, 3$, etc.

The wave will appear approximately as represented in the diagram (Fig. 7).

The equation $\varphi = \text{const.}$ represents the phase wave

[1] See, for example, *A Treatise on Light.* R. A. Houstoun. 5th edn., p. 173.

and in the same way $\overline{\frac{\partial\varphi}{\partial W}} = \text{const.}$ represents the onward motion of the maximum values we have been considering. These move forward with the group velocity.

The particular equation $\overline{\frac{\partial\varphi}{\partial W}} = 0$, the constant having in this case the particular value zero, represents the onward motion of the greatest of the series of turning values, i.e. the motion of the energy maximum.

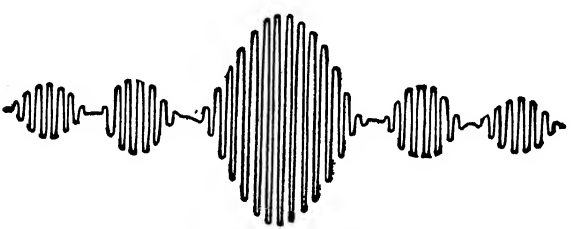


FIG. 7.

The equation $\frac{\partial\varphi}{\partial W} = 0$ may also be written thus:

$$\frac{\partial S}{\partial W} = t - t_0,$$

which is the same as the third of the group of equations (1·11), t being the time corresponding to the end of the track and t_0 that for the beginning. These two equations, which are thus actually the same, represent the motion of a particle, the former describing it as an energy node of a group of waves and the latter describing it as the motion of a massive particle by the classical method.

The Interpretation of the Quantum Conditions by Means of the Wave Theory

The fundamental quantum conditions $W = h\nu$ and $\oint p dq = nh$ have a simple interpretation in the wave theory.

This has been explained in the case of $W = h\nu$ by regarding W and $h\nu$ as the physical interpretation of a quantity which must be recognised throughout space, but which is recognised practically either as the energy of the particle or through the frequency of the associated wave.

The integrals in the second condition are to be taken over a complete circuit of values of q, p being the corresponding momentum.

Let us now interpret p by means of the phase wave, dq being along the line of propagation.

We shall take the case where the variables are separable, i.e. the function S consists of a sum of a number of functions each of which contains one variable only.

In the case where these are x, y, z we have

$$S = S_1(x) + S_2(y) + S_3(z),$$

where $S_1(x)$ depends on x only, $S_2(y)$ on y only and $S_3(z)$ on z only.

If we apply the integral theorem to the x-co-ordinate we have

$$p = G_x \text{ and } q = x,$$

hence

$$\oint G_x dx = nh.$$

Suppose that a change occurs in x, y and z remaining constant, and let x undergo the small variation from x to $x + \delta x$.

Consider a surface $S^* = \text{const.}$ or

$$S - Wt = \text{const.}$$

For the variation we are considering,

$$\frac{\partial S_1}{\partial x}\delta x - W\delta t = 0, \quad . \quad . \quad (2{\cdot}15)$$

since S_1 is the only part of S which contains x.

Thus $G_x = \frac{\partial S_1}{\partial x} = \frac{W}{U_x}$, where U_x denotes the velocity of the phase wave in the x-direction.

The quantum condition thus becomes

$$\oint \frac{W}{U_x} dx = nh$$

or

$$\oint \frac{\nu}{U_x} dx = \oint \frac{dx}{\lambda_x} = n,$$

where λ_x is the distance between places of equal phase along the x-direction. The integral gives the number of these wave-lengths in the path and the quantum condition is that a whole number of waves exist in the permissible quantum path. They are thus in the condition to persist and are not annulled by interference.

The Wave Equation

We have seen that the equation $S^* = \text{const.}$ can be regarded as representing a spreading wave surface, and we now pass on to consider the wave equation in the differential form.

In the wave theory of light we are familiar with the equation of the wave represented by (2·4). If U denote the phase velocity the equation is

$$\frac{\partial^2 \psi}{\partial x^2} = \frac{1}{U^2}\frac{\partial^2 \psi}{\partial t^2}, \quad . \quad . \quad (2{\cdot}16)$$

where ψ has been written in the place of y, and (2·4) is a solution of (2·16).

The general wave equation in three dimensions is

$$\frac{\partial^2 \psi}{\partial x^2} + \frac{\partial^2 \psi}{\partial y^2} + \frac{\partial^2 \psi}{\partial z^2} = \frac{1}{U^2}\frac{\partial^2 \psi}{\partial t^2}. \quad . \quad (2{\cdot}17)$$

and the solution of this is

$$\psi = a \sin 2\pi\left(\nu t - \frac{lx + my + nz}{\lambda} + \alpha\right). \quad (2{\cdot}18)$$

where (l, m, n) are the direction cosines of the normal to the plane wave (2·18). The expression on the left of (2·17) is denoted by $\nabla^2\psi$ by English writers and by $\Delta\psi$ by writers on the continent.

The equation is often expressed in another notation, to understand which we must refer to the meanings of the terms divergence and gradient. The divergence of a vector A which has components $(A_x\ A_y\ A_z)$ is a scalar quantity defined thus:

$$\text{div A} = \frac{\partial A_x}{\partial x} + \frac{\partial A_y}{\partial y} + \frac{\partial A_z}{\partial z}.$$

The gradient of a scalar quantity, ψ, is a vector with components, $\left(\frac{\partial\psi}{\partial x}, \frac{\partial\psi}{\partial y}, \frac{\partial\psi}{\partial z}\right)$, thus the left side of (2·17) may also be written div (grad ψ) and the wave equation (2·17) is written alternatively

$$\nabla^2\psi = \Delta\psi = \frac{1}{U^2}\frac{\partial^2\psi}{\partial t^2}$$

or

$$\text{div (grad } \psi) = \frac{1}{U^2}\frac{\partial^2\psi}{\partial t^2} \quad . \quad . \quad (2{\cdot}19)$$

The form (2·19) is particularly useful because we can use it very conveniently in generalising the wave equation for the purpose of applying it in the general theory of relativity, and in the case of mechanical systems. We must apply a similar equation to our mechanical waves, and all that is necessary is to introduce the appropriate phase velocity.

The mechanical wave equation is thus

$$\text{div (grad } \psi) = \frac{G^2}{W^2}\frac{\partial^2\psi}{\partial t^2} \quad . \quad . \quad (2{\cdot}20)$$

since by (2·3) $U = \frac{W}{G}$.

In the case of waves already familiar to the physicist ψ denotes some condition of the medium, such as displacement of its particles from a mean position or such as pressure, but the meaning of ψ in wave mechanics is not so easily interpreted. We shall later refer to the suggestions made with regard to the physical meaning of ψ.

The equation of a simple sine form of mechanical wave is

$$\psi = \psi_0 \sin \frac{2\pi}{h}(S - Wt + Wt_0), \quad . \quad (2{\cdot}21)$$

in which we have replaced the familiar phase by S* or $S - Wt + Wt_0$, which we have seen is the mechanical analogy to the phase, and where ψ_0 does not contain the time. In this case we have

$$\frac{\partial^2\psi}{\partial t^2} = -\frac{4\pi^2}{h^2}W^2\psi$$

so that (2·20) is reduced to

$$\text{div (grad } \psi) + \frac{4\pi^2}{h^2}G^2\psi = 0 \quad . \quad . \quad (2{\cdot}22)$$

From a calculation outlined earlier in this chapter (page 34) we see that

$$G^2 = m_0{}^2c^2\left\{\left(\frac{W - V}{m_0c^2}\right)^2 - 1\right\}. \quad . \quad (2{\cdot}23)$$

and we must examine the form taken by (2·22) when we wish to deal only with the equation offered by classical mechanics.

In this case the kinetic energy is $\frac{1}{2}m_0v^2$, where m_0 denotes the Newtonian mass, and the energy, E, is given by $\frac{1}{2}m_0v^2 + V$.

E is, however, an approximation to the relativity value

$$(m - m_0)c^2 + V$$

and this differs from the classical value by a quantity of the order $\left(\frac{v}{c}\right)^2$ which is very small in all classical problems.

We thus write

$$E = (m - m_0)c^2 + V$$

and

$$W = mc^2 + V = E + m_0c^2.$$

The value of G^2 may now be written

$$G^2 = \frac{(E - V)^2}{c^2} + 2m_0(E - V) \quad . \quad (2{\cdot}24)$$

Thus in any example contemplated in the classical theory we need retain only the second term on the right of (2·24), and in this term E may be given the usual value $(\frac{1}{2}m_0v^2 + V)$ without the introduction of any serious inaccuracy.

The classical wave equation is thus

$$\text{div (grad } \psi) + \frac{8\pi^2}{h^2}m_0(E - V)\psi = 0 \quad . \quad (2{\cdot}25)$$

A simple case where we have only one variable, x, is

$$\frac{\partial^2\psi}{\partial x^2} + \frac{8\pi^2}{h^2}m_0(E - V)\psi = 0$$

and by means of this equation we shall illustrate a very useful rule for obtaining the wave equation.

We write down the energy equation for this problem

$$\tfrac{1}{2}m_0\dot{x}^2 + V = E$$

and express it in terms of the momentum $p = m_0\dot{x}$

$$\frac{1}{2m_0}p^2 + V = E.$$

The Hamilton-Jacobi equation is

$$\frac{1}{2m_0}\left(\frac{\partial F}{\partial x}\right)^2 + V = E \text{ (compare 1·17[1]).}$$

We may derive the wave equation by replacing p in the energy equation or $\frac{\partial F}{\partial x}$ in the H.-J. equation by the operator $\frac{h}{2\pi i}\frac{\partial}{\partial x}$, where $i = \sqrt{-1}$.

We then obtain an operator

$$\frac{1}{2m_0}\left(\frac{h}{2\pi i}\frac{\partial}{\partial x}\right)^2 + V,$$

which is taken to be equal to the operator E, and so by allowing the two to operate on a function ψ we have

$$\left\{\frac{1}{2m_0}\left(\frac{h}{2\pi i}\cdot\frac{\partial}{\partial x}\right)^2 + V\right\}\psi = E\psi$$

or
$$\frac{h^2}{8\pi^2 m_0}\frac{\partial^2\psi}{\partial x^2} + (E - V)\psi = 0,$$

i.e.
$$\frac{\partial^2\psi}{\partial x^2} + \frac{8\pi^2 m_0}{h^2}(E - V)\psi = 0.$$

This will appear a very artificial way of arriving at the wave equation, but it will be convenient to keep it in mind.

If more than one variable be present each component of momentum, p_1, p_2, p_3, must be replaced by the corresponding operators $\frac{h}{2\pi i}\frac{\partial}{\partial x}$, $\frac{h}{2\pi i}\frac{\partial}{\partial y}$, $\frac{h}{2\pi i}\frac{\partial}{\partial z}$, and the operation upon ψ must be performed.

The introduction of the wave equation is a new departure in mechanics, the conditions involved in it give additional limitations which were not foreseen in the older classical theory. The first concern is not the meaning of the quantity ψ, but rather the limitations imposed upon E by the equation and certain conditions regarding its solution. With regard to the older quantum theory it used to be said that the quantum conditions represented certain extra limitations not included in the classical theory which were like boundary condi-

tions; this point is very clearly brought to light in the new theory.

It will be seen that the methods employed do no violence to classical dogma, they add to it, but the problem is attacked on strictly classical lines.

The Mechanical Wave-Length and Atomic Dimensions

The introduction of the constant, h, in equation (2·6) was necessary from a consideration of the dimensions of the mechanical phase. There is nothing in the theory to indicate its magnitude and this indication could not be expected. The magnitude can only be obtained by experiment. We know that this equation has been widely applied and has been experimentally verified. We are thus led to identify this constant as Planck's constant of action.

From (2·3) and from the discussion on page 44 on the magnitude of G in classical mechanics we have

$$U = \frac{W}{\sqrt{2m(E - V)}} = \frac{h\nu}{\sqrt{2m(E - V)}} \quad . \quad (2·26)$$

from which

$$\lambda = \frac{U}{\nu} = \frac{h}{\sqrt{2m(E - V)}} \quad . \quad . \quad (2·27)$$

If this be applied to the case of an electron moving about a nucleus as in the hydrogen atom, we have, writing $E - V = \frac{1}{2}mv^2$

$$\lambda = \frac{h}{mv} \quad . \quad . \quad . \quad (2·28)$$

The moment of momentum is mva, and it has the value $\frac{h}{2\pi}$ for the innermost orbit according to Bohr's theory.

Thus

$$a = \frac{h}{2\pi mv} \quad . \quad . \quad . \quad (2·29)$$

It is remarkable that the wave-length is of the same order of magnitude as a, i.e. as the atomic dimensions. We may expect from our analogy that just as the theory of diffraction comes to the aid of geometrical optics when the dimensions of the objects considered are of the same order as the wave-length of light, so the new mechanics must be invoked to aid classical mechanics when the dimensions are of the order of the mechanical wave-length which, as we have just seen, is of atomic dimensions. It is significant that our difficulties have arisen in trying to apply classical mechanics to atoms.

The Elimination of the Energy Parameter E from the Wave Equation

The equation (2·25) has acquired its particular form on account of the substitution (2·21) into (2·16). We shall now write

$$\psi = \psi_0 e^{\frac{2\pi i}{h}Wt}, \quad . \quad . \quad . \quad (2\cdot30)$$

where ψ_0 is a function of the space co-ordinates. The substitution of this value of ψ in (2·20) will again give (2·22).

This leads in the special case of classical mechanics to (2·25), which has the disadvantage that it contains the parameter, E, a constant for a particular state of the system, but varying from one state to another. We wish to obtain an equation which does not contain this constant.

For the sake of brevity we will take the case of a single space variable x.

Remembering that $W = E + m_0c^2$, we can deduce from (2·25)

$$\frac{\partial^2\psi}{\partial x^2} - \frac{4\pi i m_0 \dot{\psi}}{h} - \frac{8\pi^2 m_0}{h^2}(V + m_0c^2)\psi = 0 \quad (2\cdot31)$$

We must regard this as the fundamental equation of wave mechanics, and ψ is clearly a complex quantity.

The complex conjugate of ψ, viz. ψ^*, obtained by changing the sign of $\sqrt{-1}$ in ψ, satisfies the equation

$$\frac{\partial^2\psi^*}{\partial x^2} + \frac{4\pi i m_0}{h}\dot{\psi}^* - \frac{8\pi^2 m_0}{h^2}(V + m_0c^2)\psi = 0 \quad (2{\cdot}32)$$

This process has resulted in an equation in which the time has been restored. The equation is of the type of a diffusion equation, and can be applied to cases where V is a function of the time. It is of importance in the application of the theory to perturbations.

Thus (2·25) is to be regarded, not as the real wave equation, but as an equation in ψ_0, the amplitude, and we may describe it as an amplitude equation.

CHAPTER III

THE NATURE OF THE PROBLEM AND ITS SOLUTION

THE nature of the new problem which mechanics in this form brings before us is best understood by referring to cases of a similar character, of which other problems in physics afford numerous examples.

We will take the problem of the vibrating string, which may present itself as follows. At a given instant which we choose as the origin of our measurement of time the string is at rest in some stated form. Let us suppose it is stretched between two points at distance l apart and let the displacement, y, of the string be referred to the line joining these points which will be taken as the x-axis.

The string, which is under tension, is then released from its original position and allowed to vibrate, and we require to find the displacement at any point, distant x from one of the fixed points, at a given time t.

The wave motion of the string is represented by the equation

$$\frac{\partial^2 y}{\partial t^2} = \mathrm{U}^2 \frac{\partial^2 y}{\partial x^2} \quad . \quad . \quad . \quad (3{\cdot}1)$$

This is the wave equation in this case, and the possible solutions are infinite in number, for we know that the function

$$y = \mathrm{A}e^{ax+bt}$$

is a solution and there is no restriction on A and only one relation between a and b, viz.: $b^2 = \mathrm{U}^2 a^2$.

Nevertheless, we can obtain a definite solution because

of other conditions of the problem. These are the boundary conditions.

At all values of t the displacements at $x = 0$ and at $x = l$ are zero, so also are the velocities $\frac{dy}{dt}$ at these points; and at $t = 0$ we know the value of y for all points on the string since its original deformation is given.

So far we have a wave equation in mechanics corresponding to (3·1), but we have no other conditions. The conditions imposed are that the values of ψ must be everywhere finite, single valued and continuous, and they must approach the value zero at infinity.

The values of ψ which satisfy these conditions are described as the characteristic functions, and we shall see that only certain values may be assumed by E corresponding to these functions; these values are described as the characteristic values. Thus the problem is to find solutions of the wave equation which satisfy these conditions.

The solutions, and especially the values to which E is limited by these conditions, are of the greatest interest, and it will be worth while to consider how they are obtained in the most important examples.

It will be seen that there is no especial mathematical difficulty in deriving them, and we shall begin with a discussion of the type of equation which generally occurs in atomic problems and obtain its solution.

The wave equation in these examples presents itself ultimately in the form

$$\frac{d^2y}{dx^2} + \mathrm{P}\frac{dy}{dx} + \mathrm{Q}y = 0, \quad . \quad . \quad (3\cdot2)$$

where P and Q are polynomials in x.

This may be transformed into an equation which may be solved by means of a series, in fact we shall see that the equation may be put into the form

$$(a_2x^{k+2} + b_2x^{l+2})\frac{d^2\mathrm{X}}{dx^2} + (a_1x^{k+1} + b_1x^{l+1})\frac{d\mathrm{X}}{dx} + (a_0x^k + b_0x^l)\mathrm{X} = 0, \quad (3\cdot3)$$

where k and l are whole numbers, the particular feature of this equation being that the difference between the powers of x is the same in each bracket.

The second of these equations may be written

$$x^k\left(a_2x^2\frac{d^2\mathrm{X}}{dx^2}+a_1x\frac{d\mathrm{X}}{dx}+a_0\mathrm{X}\right)+ x^l\left(b_2x^2\frac{d^2\mathrm{X}}{dx^2}+b_1x\frac{d\mathrm{X}}{dx}+b_0\mathrm{X}\right)=0 \quad (3\cdot4)$$

By regarding $x\frac{d}{dx}$ as an operator we can express (3·4) still more conveniently if we examine the double application of this operator, which we write $\left(x\frac{d}{dx}\right)^2$. Suppose the operation is made upon a function F, then the double application on F is written $\left(x\frac{d}{dx}\right)^2\mathrm{F}$ and we have

$$\left(x\frac{d}{dx}\right)\left(x\frac{d\mathrm{F}}{dx}\right)=x^2\frac{d^2\mathrm{F}}{dx^2}+x\frac{d\mathrm{F}}{dx},$$

or
$$x^2\frac{d^2\mathrm{F}}{dx^2}=\left(x\frac{d}{dx}\right)^2\mathrm{F}-\left(x\frac{d}{dx}\right)\mathrm{F}.$$

Thus the first bracket in (3·4) may be written

$$\left\{a_2\left(x\frac{d}{dx}\right)^2+(a_1-a_2)x\frac{d}{dx}+a_0\right\}\mathrm{X}.$$

We may define a function

$$f(z)=a_2z^2+(a_1-a_2)z+a_0$$

and also
$$g(z)=b_2z^2+(b_1-b_2)z+b_0$$

and write (3·4) in the form

$$x^kf\left(x\frac{d}{dx}\right)\mathrm{X}+x^lg\left(x\frac{d}{dx}\right)\mathrm{X}=0 \quad . \quad (3\cdot5)$$

Divide throughout by x^k, it being supposed that $k>l$, and we obtain

$$f\left(x\frac{d}{dx}\right)\mathrm{X}+\frac{1}{x^c}g\left(x\frac{d}{dx}\right)\mathrm{X}=0 \; . \quad . \quad (3\cdot6)$$

where c is a positive integer.

If $l > k$ we divide by x^l and obtain an equation of the same form.

The solution of (3·6) may be solved immediately by the method of integration by series (Forsyth's *Diff. Eqns.*, ch. v), and we reproduce here as much as is necessary for our purpose in order to develop completely some of the most interesting results of the theory.

Let it be assumed that

$$X = A_1 x^{m_1} + A_2 x^{m_2} + A_3 x^{m_3} + \text{etc.},$$

where the m's are integral numbers arranged in ascending order and the A's are constants between which certain relations exist.

The method of solution consists in substituting the value of X in (3·6) and equating to zero the coefficients of the powers of x. In this way the m's and the relations between the A's are determined, and we are left finally with a solution containing the proper number of arbitrary constants, which in our case is two.

The substitution of X will give rise to terms of the form $\left(x\frac{d}{dx}\right)^2 x^n$, where n is an integer, and this is equal to $n^2 x^n$ so that after the operation we find that $x\frac{d}{dx}$ is substituted by the index of the power of x.

Thus $f\left(x\frac{d}{dx}\right)x^n = f(n)x^n$ and on substituting the series in (3·6) we find that

$$\begin{aligned} &A_1 f(m_1)x^{m_1} + A_2 f(m_2)x^{m_2} + \ldots \\ &+ A_1 g(m_1)x^{m_1 - c} + A_2 g(m_2)x^{m_2 - c} + \ldots = 0 \quad (3{\cdot}7) \end{aligned}$$

The lowest power of x in this equation is $(m_1 - c)$, and since the left-hand side must be identically zero the coefficient of $x^{m_1 - c}$ must vanish.

Hence $g(m_1) = 0$, for A_1 is not zero since it is assumed to be the coefficient of a term present in the series.

The roots of this equation give the values of m_1 with which the series begins, and in the case we are considering $g(z)$ is a quadratic and the number of roots is two.

On examining the other terms of the series we find that a solution can be obtained in this way by equating the indices thus :

$$m_1 = m_2 - c$$
$$m_2 = m_3 - c, \text{ etc.}$$

and this is possible since they are arranged in ascending order. From this it follows that the series advances in powers of x differing by c.

Again, since the terms vanish identically, we must have

$$A_1 f(m_1) + A_2 g(m_2) = 0$$
$$A_2 f(m_2) + A_3 g(m_3) = 0, \text{ etc.},$$

from which all the coefficients may be obtained in terms of the first.

Suppose that a is one of the values of m_1 with which the series begins, then we have

$$A_2 = -\frac{f(a)}{g(a+c)}A_1, A_3 = \frac{f(a+c)f(a)}{g(a+2c)\,g(a+c)}A_1, \text{ etc.},$$

and the solution is

$$X = A_1 x^a \left\{1 - \frac{f(a)}{g(a+c)}x^c + \frac{f(a+c)f(a)}{g(a+2c)\,g(a+c)}x^{2c} - \text{etc.}\right\} . \quad (3\cdot8)$$

There are as many series as there are roots of the equation

$$g(m_1) = 0$$

and in the case in which we are interested this number is two, and the complete solution will be the sum of (3·8) and another series like it with another arbitrary constant instead of A_1 and beginning with another power of x than a.

Certain complications may occur in the series, for we may have a zero factor in the numerator, denominator, or in both simultaneously in the coefficients of the series. In some cases the two series are infinite.

We have no need to go into these cases in detail, for provided that the series are convergent we have only two points to bear in mind; the first is that if one of the series terminates and therefore contains only a finite number of terms this corresponds to the occurrence of a line spectrum and the appropriate values of E are the corresponding energy levels.

If the series is infinite this corresponds to a continuous spectrum.

The condition that the series shall terminate is that $f(a)$, $f(a + c)$ or one of the functions occurring in the numerator shall vanish.

We will first examine the condition for convergence.

A series $U_1 + U_2 + U_3 + \ldots$ is convergent provided that after some term in the series the ratio of one term to the next is greater than unity, i.e. $\left|\frac{U_n}{U_{n+1}}\right| > 1$ for values of n greater than some particular value.

If we apply this to the series (3·8) we find that the condition is

$$\left|\frac{g(a + \overline{n+1}c)}{f(a + nc)} \cdot \frac{1}{x^c}\right| > 1.$$

Thus remembering the form of g and f we find, when n is large, that

$$\left|\frac{b_2}{a_2} \cdot \frac{1}{x^c}\right| > 1, \text{ or } x^c < \left|\frac{b_2}{a_2}\right|.$$

It will be noted that if a_2 is zero, the term a_2x^{k+2} not occurring in the equation, the series is convergent for all finite values of x no matter how large.

In the cases we have to consider the term a_2x^{k+2} does not occur, so that no difficulty arises with regard to convergence.

The equation (3·3) can now be written thus:

$$\frac{d^2X}{dx^2} + \left(a_1x^{s+1} + \frac{b_1}{x}\right)\frac{dX}{dx} + \left(a_0x^s + \frac{b_0}{x^2}\right)X = 0 \quad (3·9)$$

where slight changes have been made in the use of the constants, s being written for $(k - l - 2)$.

We have considered the method of solution of (3·9) which we have stated is the form on which the solution of the wave equation ultimately depends. We must consider how the wave equation can be reduced to this known form.

We begin in general with (3·2) and make the substitution $y = \mathrm{FX}$, and since X is a finite series we require that F should approach zero as x approaches infinity in order to satisfy our boundary condition.

On substituting for y we obtain

$$\frac{d^2\mathrm{X}}{dx^2} + \left(\frac{2}{\mathrm{F}}\frac{d\mathrm{F}}{dx} + \mathrm{P}\right)\frac{d\mathrm{X}}{dx} + \left(\frac{1}{\mathrm{F}}\frac{d^2\mathrm{F}}{dx^2} + \frac{\mathrm{P}}{\mathrm{F}}\frac{d\mathrm{F}}{dx} + \mathrm{Q}\right)\mathrm{X} = 0 \quad (3\cdot10)$$

In order that this shall be of the form (3·9) we must choose F to satisfy the conditions

$$\frac{2\mathrm{F}'}{\mathrm{F}} + \mathrm{P} = a_1 x^{s+1} + \frac{b_1}{x},$$

$$\frac{\mathrm{F}''}{\mathrm{F}} + \frac{\mathrm{P}\mathrm{F}''}{\mathrm{F}} + \mathrm{Q} = a_0 x^s + \frac{b_0}{x^2},$$

writing F′ and F″ for $\frac{d\mathrm{F}}{dx}$ and $\frac{d^2\mathrm{F}}{dx^2}$ respectively.

Now P and Q are polynomials, so that $\frac{\mathrm{F}'}{\mathrm{F}}$ and $\frac{\mathrm{F}''}{\mathrm{F}}$ are also polynomials, and consequently the differential coefficients must be divisible by F. We write, therefore, $\mathrm{F} = e^{\phi}$ and have to choose ϕ to satisfy

$$\left.\begin{aligned} 2\phi' + \mathrm{P} &= a_1 x^{s+1} + \frac{b_1}{x} \\ \phi'' + \phi'^2 + \mathrm{P}\phi' + \mathrm{Q} &= a_0 x^s + \frac{b_0}{x^2} \end{aligned}\right\} . \quad (3\cdot11)$$

ϕ' is thus a polynomial in x, and if ϕ is a polynomial function of x with negative coefficients e^{ϕ} will approach zero as x approaches infinity, and our condition at infinity is satisfied.

We need not proceed farther to express F in terms of x in the general case, we shall have examples of this in the cases we have to consider.

The Planck Oscillator

The first example is one of historical interest because it is that to which Planck first applied his original quantum postulate.

It is an excellent example of the application of the principles of the new mechanics.

The oscillator is a particle vibrating under a restoring force proportional to the displacement and executes simple harmonic vibrations.

Suppose that the particle vibrates along the axis of x and that at any instant it is displaced to a point distant x from this mean position. If the restoring force is denoted by kx the equation of motion is

$$m\ddot{x} = -kx \quad . \quad . \quad . \quad (3{\cdot}12)$$

The solution of this equation is

$$x = \mathrm{A} \sin (2\pi\nu t + \alpha),$$

where A and α are constants and $4\pi^2\nu^2 = \frac{k}{m}$, ν denoting the frequency of the vibrations.

The potential energy at x is $\frac{1}{2}kx^2 = 2m\pi^2\nu^2x^2$.

The energy equation is

$$\tfrac{1}{2}m\dot{x}^2 + \tfrac{1}{2}kx^2 = \mathrm{E}$$

and writing $p = m\dot{x}$ we have

$$\frac{1}{2m}p^2 + 2m\pi^2\nu^2x^2 = \mathrm{E}.$$

Thus according to the rule given at the end of the last chapter the wave equation is

$$\frac{1}{2m}\left(\frac{h}{2\pi i}\frac{d}{dx}\right)^2\psi + 2m\pi^2\nu^2x^2\psi = \mathrm{E}\psi$$

or

$$\frac{d^2\psi}{dx^2} + \frac{8\pi^2m}{h^2}(\mathrm{E} - 2m\pi^2\nu^2x^2)\psi = 0.$$

If we write $a = \frac{8\pi^2 mE}{h^2}$ and $b = \frac{16\pi^4 m^2 \nu^2}{h^2}$ for convenience, this equation is more simply written

$$\frac{d^2\psi}{dx^2} + (a - bx^2)\psi = 0 \quad . \quad . \quad (3\cdot13)$$

This equation is of the type (3·2), and we can convert it to the type we have been considering by means of equations (3·11).

In this special case $P = 0$, $Q = a - bx^2$, and it will be noted that if in (3·11) we write

$$s = 0,\ b_1 = 0,\ a_1 = -2b^{\frac{1}{2}}$$

we have $\phi = -\frac{1}{2}b^{\frac{1}{2}}x^2$. The negative sign is required to make ψ remain finite and approach zero when x approaches infinity.

The second equation in (3·11) may be satisfied by choosing the right values for a_0 and b_0, which are

$$a_0 = a - b^{\frac{1}{2}} \text{ and } b_0 = 0.$$

It is simplest to obtain the transformed equation by making the substitution directly in (3·13) rather than to use the general case; thus on substituting $\psi = e^{-\frac{1}{2}b^{\frac{1}{2}}x^2}\mathrm{X}$ we find

$$\frac{d^2\mathrm{X}}{dx^2} - 2b^{\frac{1}{2}}x\frac{d\mathrm{X}}{dx} + (a - b^{\frac{1}{2}})\mathrm{X} = 0 \quad . \quad (3\cdot14)$$

which is of the form (3·9), and of which we have studied the solution.

For the sake of convenience we write (3·14) thus:

$$\frac{d^2\mathrm{X}}{dx^2} - cx\frac{d\mathrm{X}}{dx} + e\mathrm{X} = 0$$

and make the substitution

$$\mathrm{X} = \Sigma \mathrm{A}_1 x^{m_1}.$$

It then follows that the values of the lowest index m_1 are either zero or unity, and that each series advances in powers increasing by 2.

The constant coefficients are given in terms of the lowest by

$$A_2 = \frac{(cm_1 - e)}{(m_1 + 2)(m_1 + 1)} A_1,$$

$$A_3 = \frac{(cm_1 + 2 - e)(cm_1 - e)}{(m_1 + 4)(m_1 + 3)(m_1 + 2)(m_1 + 1)} A_1, \text{ etc.}$$

There is no difficulty in either of the series which arise when the values $m_1 = 0$ and $m_1 = 1$ are introduced, for no zero factor occurs anywhere in the denominators.

We have to decide whether either or both of the series will terminate, and this, as we have mentioned, corresponds to the question in physics as to whether this oscillator will give rise to line spectra or continuous spectra or to one of each.

The series will terminate if one of the factors in the numerator vanishes, for if in advancing in the series any A vanishes then all succeeding A's vanish by the above relations.

Now the factors in the numerator are of the form

$$c(m_1 + 2n) - e,$$

where n is an integer.

This will vanish if

$$n = \frac{e - m_1 c}{2c},$$

i.e. if the quantity on the right is an integer.

If we take the case when $m_1 = 0$ we see that this requires $\frac{e}{2c}$ to be an integer.

Now if we refer to the quantities for which e and c stand, we find that

$$\frac{e}{c} = \frac{a}{2b^{\frac{1}{2}}} - \tfrac{1}{2}$$

$$= \frac{E}{h\nu} - \tfrac{1}{2}.$$

Thus the series terminates if $E = (2n + \frac{1}{2})h\nu$.

Again, if $m_1 = 1$ the series will terminate if

$$c(1+2n) - e = 0,$$

i.e. if $$2n + 1 = \frac{e}{c} = \frac{E}{h\nu} - \tfrac{1}{2},$$

i.e. if $$E = (2n + \tfrac{3}{2})h\nu$$
$$= (2n + 1 + \tfrac{1}{2})h\nu.$$

From the two results the possible values of E may be written

$$E = (n + \tfrac{1}{2})h\nu$$

where n denotes any integer, odd or even.

These values of E may be said to be those physically possible. There is nothing in our derivation to show that others may not occur, but apparently this is the case, for it is well known that the assumption that the oscillator possesses these particular energy values has cleared up a very great difficulty in the problem of radiation.

The occurrence of the $\frac{1}{2}h\nu$ is noteworthy, for the older view was that $E = nh\nu$. The formula obtained here is identical with that obtained by Heisenberg's method.

The Rotator

In the next example we take the case of a body rotating about a fixed axis. Rotators of this type are considered in the theory of band spectra, although in this case the problem is not so simple as that we shall consider, for the axis in our case is supposed to be fixed in space, while in the band-spectrum problem it is free.

Our example has no physical application, but it is the simplest of all wave mechanics problems and it illustrates a case in which it is difficult to form any picture of actual waves. It is, in fact, a generalised dynamical wave problem with one degree of freedom.

There is no potential energy, and if I denote the

moment of inertia and θ the angular co-ordinate, have for the energy equation simply

$$\tfrac{1}{2}\mathrm{I}\dot{\theta}^2 = \mathrm{E}.$$

Writing $p = \dfrac{\partial \mathrm{T}}{\partial \dot{\theta}} = \mathrm{I}\dot{\theta}$ we deduce

$$\frac{1}{2\mathrm{I}}p^2 = \mathrm{E}.$$

Thus the wave equation is

$$\frac{1}{2\mathrm{I}}\left(\frac{h}{2\pi i}\frac{d}{d\theta}\right)^2\psi = \mathrm{E}\psi$$

or

$$\frac{d^2\psi}{d\theta^2} + \frac{8\pi^2\mathrm{I}\mathrm{E}\psi}{h^2} = 0.$$

In this case the momentum p and the co-ordinate correspond, and the operator to be substituted for p

$$\frac{h}{2\pi i}\frac{\partial}{\partial\theta} \text{ (p. 45).}$$

The solution is

$$\psi = \mathrm{A}\sin(k\theta + \alpha),$$

where $k^2 = \dfrac{8\pi^2\mathrm{I}\mathrm{E}}{h^2}$, and A and α are constants.

If ψ is to be single valued it must have the same value when θ changes by 2π, and consequently it is periodic in θ.

Thus k is an integer, n, and we find

$$\mathrm{E} = \frac{h^2}{8\pi^2\mathrm{I}}n^2.$$

The form of this simple result is very suggestive, for it is a reminder of the band-spectra energy formulæ, but on considering the case for the moving axis the energy values are

$$\mathrm{E} = \frac{n(n+1)}{8\pi^2\mathrm{I}}h^2.$$

This is a surprising result, for not only does it give the actual energy values, but it gives immediately the half-quantum numbers which have been most difficult to explain in this connection.

The formula for E can be written

$$\mathrm{E} = \frac{h^2}{8\pi^2 \mathrm{I}}\{(n + \tfrac{1}{2})^2 - \tfrac{1}{4}\}$$

and the differences corresponding to two energy levels are thus of the form

$$\mathrm{E}' - \mathrm{E}'' = \frac{h^2}{8\pi^2 \mathrm{I}}\{(n' + \tfrac{1}{2})^2 - (n'' + \tfrac{1}{2})^2\}.$$

The Line Spectrum of Hydrogen

Another problem of importance is given by the hydrogen atom, and we now undertake the determination of the energy levels in this case. This is treated in the same way as the others, the only difference being that the details of the calculations are a little longer.

The mechanical system in this case consists of a central positive charge e, and an electronic charge of equal magnitude at a distance r from it.

We may derive the wave equation by the symbolical method we have just employed, but we will do so directly in this case in order to illustrate the other way of determining the equation in a special problem.

In the equation (2·25) we have to substitute $\mathrm{V} = -\frac{e^2}{r}$ in order to deal with our problem, and the wave equation is consequently

$$\frac{\partial^2\psi}{\partial x^2} + \frac{\partial^2\psi}{\partial y^2} + \frac{\partial^2\psi}{\partial z^2} + \frac{8\pi^2 m}{h^2}\left(\mathrm{E} + \frac{e^2}{r}\right)\psi = 0 . \quad (3\cdot15)$$

On account of the occurrence of r we shall make use of polar co-ordinates. The first three terms then assume a well-known form which we shall write down without

proof since the transformation is readily made and is given in mathematical text-books.

The transformed equation is

$$\frac{1}{r^2}\frac{\partial}{\partial r}\left(r^2\frac{\partial\psi}{\partial r}\right)+\frac{1}{r^2\sin\theta}\frac{\partial}{\partial\theta}\left(\sin\theta\frac{\partial\psi}{\partial\theta}\right)+\frac{1}{r^2\sin^2\theta}\frac{\partial^2\psi}{\partial\phi^2}+\frac{8\pi^2m}{h^2}\left(\mathrm{E}+\frac{e^2}{r}\right)\psi=0 \quad . \quad . \quad (3\cdot16)$$

The method adopted for solving this equation is to try the substitution

$$\psi=\mathrm{R}\Theta\Phi$$

where R is a function of r only, Θ of θ only and Φ of ϕ only. (3·16) then becomes

$$\frac{1}{\mathrm{R}r^2}\frac{\partial}{\partial r}\left(r^2\frac{\partial\mathrm{R}}{\partial r}\right)+\frac{1}{\Theta r^2\sin\theta}\frac{\partial}{\partial\theta}\left(\sin\theta\frac{\partial\Theta}{\partial\theta}\right)+\frac{1}{\Phi r^2\sin^2\theta}\frac{\partial^2\Phi}{\partial\phi^2}+\frac{8\pi^2m}{h^2}\left(\mathrm{E}+\frac{e^2}{r}\right)=0.$$

From the fact that R and Θ do not contain ϕ we know that the only term that could contain this co-ordinate is $\frac{1}{\Phi}\frac{\partial^2\Phi}{\partial\phi^2}$, but the equation shows that this term does not contain it.

Thus $\frac{1}{\Phi}\frac{\partial^2\Phi}{\partial\phi^2}$ is independent of r, θ or ϕ, and is thus a constant. Let this constant be written $-p^2$.

The equation now becomes

$$\frac{1}{\mathrm{R}r^2}\frac{\partial}{\partial r}\left(r^2\frac{\partial\mathrm{R}}{\partial r}\right)+\frac{1}{r^2}\left\{\frac{1}{\Theta\sin\theta}\frac{\partial}{\partial\theta}\left(\sin\theta\frac{\partial\Theta}{\partial\theta}\right)-\frac{p^2}{\sin^2\theta}\right\}+\frac{8\pi^2m}{h^2}\left(\mathrm{E}+\frac{e^2}{r}\right)=0.$$

By the same argument as the above it follows that the quantity in brackets { } is a constant. It is usual to write this constant in the form $-n(n+1)$, where n may be any integer or may have the value zero. This limitation upon the form of the constant is necessary

in order that Θ may be a single-valued function of θ. The equation now takes the form

$$\frac{d^2R}{dr^2} + \frac{2}{r}\frac{dR}{dr} + \left\{\frac{8\pi^2 mE}{h^2} + \frac{8\pi^2 me^2}{r} - \frac{n(n+1)}{r^2}\right\}R = 0 \qquad (3\cdot16)^1$$

These considerations are an outline of a general treatment which precedes one method of beginning the study of spherical and tesseral harmonics. There has so far been nothing which is especially characteristic of our problem. The equation (3·16) is, however, peculiar to our special case in that it contains the term in $\frac{1}{r}$. It is thus from the form of R that we shall derive information of interest in this example.

The actual value of ψ is the product of R, Θ and Φ. The last of these is equal to $(A_p \sin p\phi + B_p \cos p\phi)$ where A_p and B_p are constants and p may have any integral value.

Θ is the solution of the equation obtained by equating the expression in the brackets $\{\}$ to $-n(n+1)$. It is a rather complicated function, but very well known to the mathematician and widely applied in physics. We shall not require these functions, and so pass on to find what limitations (3·16) will place upon E. This equation is of the form we have discussed, and it will be found by the above methods that the appropriate substitution is

$$R = e^{\frac{1}{2}a_1 r}X \qquad (3\cdot17)$$

and that 3·16 then gives

$$\frac{d^2X}{dr^2} + \left(a_1 + \frac{b_1}{r}\right)\frac{dX}{dr} + \left(\frac{a_0}{r} + \frac{b_0}{r^2}\right)X = 0 \qquad (3\cdot18)$$

where

$$\left.\begin{aligned} \tfrac{1}{4}a_1^2 &= -\frac{8\pi^2 mE}{h^2} \\ a_0 &= a_1 + \frac{8\pi^2 me^2}{h^2} \\ b_1 &= 2 \\ b_0 &= -n(n+1) \end{aligned}\right\} \qquad (3\cdot19)$$

The equation (3·18) may be written

$$\left(a_1 r\frac{d}{dr} + a_0\right)\mathrm{X} + \frac{1}{r}\left\{\left(r\frac{d}{dr}\right)^2 + r\frac{d}{dr} + b_0\right\}\mathrm{X} = 0 \quad (3\cdot20)$$

or writing $f(z) = a_1 z + a_0$
$g(z) = z^2 + z + b_0$

we have

$$f\left(r\frac{d}{dr}\right)\mathrm{X} + \frac{1}{r}g\left(r\frac{d}{dr}\right)\mathrm{X} = 0 \quad (3\cdot21)$$

which is of the form (3·6).

The solution is thus given by the sum of two series of the form (3·8), where the appropriate values of a are the roots of the equation

$$g(m_1) = 0$$

i.e. $$m_1^2 + m_1 + b_0 = 0.$$

On making the substitution for b_0 given by (3·19), we find that the two values are n and $-(n+1)$, so that the series begin with r^n and $r^{-(n+1)}$.

If we consider the series beginning with $r^{-(n+1)}$, we see that it will terminate if $f(a)$, $f(a+c)$, or any such term as $f(a+sc)$ vanish, where s is an integer and c is unity in our problem. Thus, since $f(z) = a_1 z + a_0$, we have to examine under what conditions

$$a_1(a + sc) + a_0$$

will become zero.

Then writing $a = -(n+1)$, $c = 1$, we find that this requires that

$$-n - 1 + s = -\frac{a_0}{a_1}$$

or $$s - n = \frac{a_1 - a_0}{a_1}.$$

By making use of the values of a_1 and a_0 given above, we find that the condition is

$$(s-n)^2 = -\frac{2\pi^2 me^4}{\mathrm{E}h^2}.$$

Thus since s and n are integers, we may denote $(s - n)$ by an integer τ and write

$$E = -\frac{2\pi^2 me^4}{h^2\tau^2}.$$

It follows that the energy levels in the line series of the hydrogen spectrum are given by this series of values of E. These are, of course, the Bohr energy values.

In order that R, and consequently ψ, may become zero at infinity, the value of a_1 in the exponential factor $e^{ia_1 r}$ must be negative and real. Thus the value of E must be negative (3·19), and the reader should compare this with Schrödinger's discussion in his first paper (4), (*Ann. d. Physik*, Vol. 79, 1926).

These values of E account for the Balmer series in a manner familiar in the old quantum theory, but no explanation is given for the fine structure of the lines in this series in the account just given.

The Rotator with a Free Axis

The problem now to be considered is that which has been mentioned in the second example, it is that of a rotator about an axis which is free. As an illustration of the system we may imagine two spheres at a fixed distance apart rotating about an axis bisecting perpendicularly the line joining their centres. We have a dumb-bell with the two spheres lying in a plane and rotating about a normal to the plane. In the former case which we considered this plane was fixed in space, i.e. the axis of rotation was fixed; we now consider the case in which the plane can move, the axis being free to take

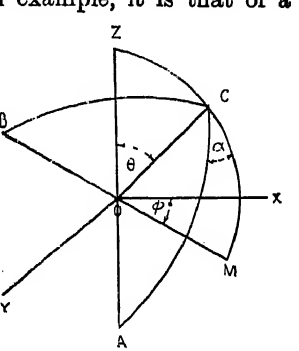


Fig. 8.

up any position. We do not contemplate any motion of the dumb-bell about the line of centres.

Let OC denote the line of centres of the system, so that it is the axis about which there is no rotation. Let the position of OC in space be described in the usual way by θ and ϕ, and let ZOCM be the plane on which the system rests.

Let OA and OB be any two lines perpendicular to each other and to OC.

For convenience we shall suppose that OA, OB and OC are of unit length, and let AC and BC be joined by circular arcs. Since OC is moving freely the system will have a rotation which may be considered as having components about OA and OB, these components being denoted by ω_1 and ω_2. It is, of course, well known that a rotation about an axis may be treated as a vector, and since there is no rotation about OC the axis of rotation must lie on the plane OAB.

Let us suppose that the system is such that the moment of inertia about any line in the plane OAB is I. Thus since the angular velocities about OA and OB are ω_1 and ω_2, the kinetic energy of the system is

$$\tfrac{1}{2}\mathrm{I}(\omega_1^2 + \omega_2^2).$$

We must express ω_1 and ω_2 in terms of θ and ϕ.

Let α denote the angle between the plane OAC and OCM, then the velocity of C along ZCM is $\dot{\theta}$ and perpendicular to the plane ZOM, $\sin\theta \,.\, \dot{\phi}$. Expressed in terms of ω_1 and ω_2 the velocity of C is ω_1 along the arc CB and ω_2 along AC, where ω_1 and ω_2 are regarded as right-handed twists about OA and OB.

Thus the velocity of C along ZCM is

$$-(\omega_1 \sin\alpha + \omega_2 \cos\alpha)$$

and perpendicular to ZOM $(\omega_1 \cos\alpha - \omega_2 \sin\alpha)$.

Thus $$\dot{\theta} = -(\omega_1 \sin\alpha + \omega_2 \cos\alpha)$$

and $$\sin\theta \,.\, \dot{\phi} = \omega_1 \cos\alpha - \omega_2 \sin\alpha.$$

From these equations we find that the above expression for the kinetic energy is

$$T = \tfrac{1}{2}I(\dot{\theta}^2 + \sin^2\theta \cdot \dot{\phi}^2)$$

or writing

$$p_1 = \frac{\partial T}{\partial \dot{\theta}},\ p_2 = \frac{\partial T}{\partial \dot{\phi}}$$

$$T = \frac{1}{2I}\left(p_1^2 + \frac{p_2^2}{\sin^2\theta}\right).$$

In this case there is no potential energy, so that

$$\frac{1}{2I}\left(p_1^2 + \frac{p_2^2}{\sin^2\theta}\right) = E.$$

This is a case in which we have two generalised co-ordinates θ and ϕ, so that our space is of two dimensions. In order to find the wave equation we must determine $\nabla^2\psi$ in this space and then express the equation

$$\nabla^2\psi + \frac{8\pi^2 m}{h^2}E\psi = 0 \quad . \quad . \quad (3{\cdot}22)$$

in θ and ϕ.

T is exactly of the form which occurs for a particle moving on the surface of unit sphere, the only difference being that I replaces the mass of the particle.

Thus the method of transforming $\nabla^2\psi$ in this case is simply to place $r = 1$ in (3·16) in the first three terms, leaving for the wave equation

$$\frac{1}{\sin\theta}\frac{\partial}{\partial\theta}\left(\sin\theta\frac{\partial\psi}{\partial\theta}\right) + \frac{1}{\sin^2\theta}\frac{\partial^2\psi}{\partial\phi^2} + \frac{8\pi^2 IE}{h^2}\psi = 0 \quad (3{\cdot}23)$$

The method of solution is similar to that employed in the problem of the hydrogen atom, and from the discussion in connection with it we see that the condition that ψ shall be single-valued in θ requires that

$$\frac{8\pi^2 IE}{h^2} = n(n+1)$$

where n is an integer.

Thus

$$E = \frac{h^2}{8\pi^2 I} n(n+1) = \frac{h^2}{8\pi^2 I}\{(n+\tfrac{1}{2})^2 - \tfrac{1}{4}\}$$

which gives the result quoted in connection w problem of the rigid rotator about the fixed axis interesting to see how the integer values and t quantum numbers come in by the application of well-known result in mathematics which we re part of the properties of the legitimate stage of physics.

The Occurrence of Half-Quantum Nume

Half-quantum numbers occur naturally in the values according to our theory and merit no consideration from this point of view. They ha a source of difficulty and the subject of much sion in the old quantum theory, and it is wort to examine the conditions under which they o are absent.

We have seen that the solution of the problem c finally on the equation

$$\frac{d^2X}{dx^2} + \left(a_1 x^{s+1} + \frac{b_1}{x}\right)\frac{dX}{dx} + \left(a_0 x^s + \frac{b_0}{x^2}\right)X = 0$$

This may be written in the form

$$f\left(x\frac{d}{dx}\right)X + \frac{1}{x^{s+2}} g\left(x\frac{d}{dx}\right)X = 0$$

where $f(z) = a_1 z + a_0$
$g(z) = z^2 + (b_1 - 1)z + b_0.$

Let α denote a root of $g(z) = 0$, and suppose

$$\alpha = \frac{1 - b_1}{2} - \tfrac{1}{2}\sqrt{(b_1 - 1)^2 - 4b_0}.$$

The series corresponding to this root begins and advances with the indices increasing by from term to term.

The series terminates if

$$f(a + nc) = 0,$$

where c is written for $(s + 2)$ and n is an integer.

This condition is

$$a_1(a + nc) + a_0 = 0$$

or
$$nc = -a - \frac{a_0}{a_1}$$
$$= -\tfrac{1}{2} + \frac{b_1}{2} - \frac{a_0}{a_1} + \tfrac{1}{2}\sqrt{(b_1 - 1)^2 - 4b_0},$$

and there is a similar condition for the other root containing a negative sign before the square root in the last equation.

The occurrence of the first term on the right is noteworthy, for it is responsible for the half-quantum number.

Whole numbers occur in the case where $b_0 = 0$, since the last term contains $\frac{1}{2}$.

This has been illustrated in the problem of the hydrogen spectrum where $c = 1$ and $b_0 = 0$ or $\neq 0$ according as we take the mass of the electron as constant or as a function of the speed.

The fraction $\frac{1}{2}$ will be reduced to $\frac{1}{4}$ in a problem in which $c = 2$, and for different values of c it will be replaced by other values.

The Theory or Radioactivity

In all the problems considered in this chapter we have found no occasion to make use of the function ψ itself. The solutions we have obtained have indicated particular energy values.

In the last example we shall consider a case in which the function itself is of importance. We shall discuss the nature of ψ in a later chapter, and shall to some extent anticipate the remarks made there.

The problem for our consideration concerns the passage of a charged particle through a potential energy barrier.

To take a particular example of a problem of this type, let us consider the potential in the neighbourhood of the nucleus of an atom. It has often been suggested that in its immediate neighbourhood the forces are attractive and not repulsive according to Coulomb's law of inverse squares, as is the case farther away. We know from experiments on the scattering of α-particles that, in the case of the heavy elements the law of repulsion holds up

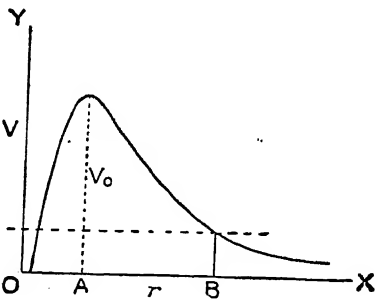


FIG. 9.

to distances of the order of 10^{-12} cm., and therefore it is only within this distance that the attractive forces become appreciable. We can thus represent the form of the potential near the nucleus by means of the curve of Fig. 9.

Let us suppose the maximum potential to be V_0 at a distance r_0 from the nucleus. Any α-particle in the nucleus has thus to overcome this potential barrier if it is to escape from the atom. The old theory of dynamics teaches that if the α-particle in the nucleus has less energy than that which would be its potential energy at the distance r_0 it cannot get away.

e its total energy is E, K kinetic, and V
n

$$E = K + V < V_0.$$

nable V to attain the value V_0 and so just
ıust assume a negative value. Negative
y is not recognised in classical dynamics,
pretation of the above inequality would be
rticle never escapes.

n to an actual example, we find that
of uranium α-particles are emitted which
mount of energy equal to their potential
distance of $6{\cdot}3 \times 10^{-12}$ cm., let us say at

This is the energy of the α-particle, and
energy when located in the nucleus. But
periments show that Coulomb's law holds
listance of $3{\cdot}2 \times 10^{-12}$ cm. of the nucleus.
icle in escaping had to cross a region where
energy was greater than its own energy.
ion of this fact was impossible on the basis
ıeory.

ve mechanics there is no difficulty. Here
t is based upon probabilities, and there is
bability greater than zero that a particle in
ergy E_0 will pass to another region of energy
ıgh the two regions may be separated by a
ter energy.

onsider a problem in which the potential
simpler than that prevailing in the region
ıs of an atom, but we shall be able to ap-
ıature and solution of the more complicated
treatment of this artificial one.

he solution of the problem presented by
processes to the work of Gamow (*Zeit. für*
51, p. 204, 1928) and of Condon and Gurney
d Sept., 1928; and *Phys. Review*, vol. 33,
). We consider three regions, 1, 2, 3. In
suppose that the potential energy of a par-
while in 2 it is V_0. We shall suppose that

the energy of the particle is E, and that $E < V_0$. The question to be answered is whether, according to the theory of wave mechanics, a particle in 1 can travel

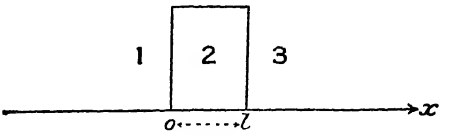


FIG. 10.

through 2 and travel away from the latter in the region 3. We know that experiment gives an affirmative answer. We have thus to find whether a solution of the wave equation exists in the form of a wave travelling along the positive direction of x in region 3. The solution will clearly be of the form

$$\psi = Ae^{2\pi i(\nu t - x/\lambda)}$$

with $\nu = \frac{W}{h}$, for this represents a wave of frequency ν travelling in this direction.

If we refer to the substitution (2·30) and to (2·25) we see that we require a solution of the equation

$$\frac{\partial^2\psi_0}{\partial x^2} + \frac{8\pi^2 m_0 E}{h^2}\psi_0 = 0 \qquad . \quad (3{\cdot}26)$$

in the form $\psi_0 = A_3 e^{-\alpha_3 x}$, since this is an amplitude equation with the factor $e^{2\pi i\nu t}$ eliminated (*cf.* 2·21).

We have to satisfy the following equations:

for region 1,

$$\frac{\partial^2\psi_0}{\partial x^2} + \alpha_1{}^2\psi_0 = 0, \ \alpha_1{}^2 = \frac{8\pi^2 m_0 E}{h^2} \qquad . \quad (3{\cdot}27)$$

for region 2,

$$\frac{\partial^2\psi_0}{\partial x^2} - \alpha_2{}^2\psi_0 = 0,\ \alpha_2{}^2 = \frac{8\pi^2 m_0(V_0 - E)}{h^2} \quad (3{\cdot}28)$$

for region 3,

$$\frac{\partial^2\psi_0}{\partial x^2} + \alpha_3{}^2\psi_0 = 0,\ \alpha_3{}^2 = \alpha_1{}^2 \quad . \quad (3{\cdot}29)$$

The solutions of these equations are :

$$\psi_{01} = A_1 e^{i\alpha_1 x} + B_1 e^{-i\alpha_1 x} \quad . \quad . \quad (3{\cdot}30)$$
$$\psi_{02} = A_2 e^{\alpha_2 x} + B_2 e^{-\alpha_2 x} \quad . \quad . \quad (3{\cdot}31)$$
$$\psi_{03} = B_3 e^{-i\alpha_3 x} \quad . \quad . \quad . \quad (3{\cdot}32)$$

In (3·32) we have omitted the term $A_3 e^{i\alpha_3 x}$, since we are concerned only with the wave along the positive direction.

We have also to satisfy the conditions of continuity :

$$\psi_{01} = \psi_{02},\quad \frac{\partial\psi_{01}}{\partial x} = \frac{\partial\psi_{02}}{\partial x} \text{ at } x = 0,$$

when we take the plane separating 1 and 2 as $x = 0$, and

$$\psi_{02} = \psi_{03},\quad \frac{\partial\psi_{02}}{\partial x} = \frac{\partial\psi_{03}}{\partial x} \text{ at } x = l,$$

where $x = l$ is the plane separating 2 and 3.

We thus obtain the following equations between the A's and B's :

$$A_1 + B_1 = A_2 + B_2$$
$$i\alpha_1(A_1 - B_1) = \beta(A_2 - B_2)$$
$$A_2 e^{\alpha_2 l} + B_2 e^{-\alpha_2 l} = B_3 e^{-i\alpha_3 l}$$
$$\beta(A_2 e^{\alpha_2 l} - B_2 e^{-\alpha_2 l}) = -i\alpha_3 B_3 e^{-i\alpha_3 l}.$$

From these we can deduce

$$A_1 = \tfrac{1}{2}B_3 i\left(\frac{\alpha_3}{\alpha_2} + \frac{\alpha_2}{\alpha_1}\right)e^{-i\alpha_3 l}\sinh\alpha_2 l$$

whence $\quad |A_1|^2 = |B_3|^2\,\tfrac{1}{4}\left(\frac{\alpha_3}{\alpha_2} + \frac{\alpha_2}{\alpha_1}\right)^2 \sinh^2\alpha_2 l$

or, after putting in the values of the α's,

$$\frac{|B_3|^2}{|A_1|^2} = 16\frac{E}{V_0}\left(1 - \frac{E}{V_0}\right)\sinh^{-2}\alpha_2 l \quad . \quad (3\cdot33)$$

Thus, if we take the case when $\alpha_2 l$ is very large, this ratio becomes

$$16\frac{E}{V_0}\left(1 - \frac{E}{V_0}\right)e^{-2\alpha_2 l} \quad . \quad . \quad (3\cdot34)$$

As we shall see in a later chapter, this ratio measures the relative probabilities of the presence of the particle in the regions 1 and 3. Another way of stating this is that given the existence of a particle in region 1 the ratio gives the probability of its escape to 3, and out of a large number of particles in 1 the ratio represents the proportion which escape.

The controlling factor of the expression, except when $\frac{E}{V_0}$ is zero or unity, is the exponential term, and we can write $2\alpha_2 l$ in a form which is useful in more general applications of the theory, viz. :

$$J = \int \frac{4\pi}{h}\sqrt{2m_0(V_0 - E)}dx \quad . \quad . \quad (3\cdot35)$$

In the present case $(V_0 - E)$ is constant across the potential barrier, but even when this is not the case the integral form holds and the value can be represented graphically by plotting $\frac{4\pi}{h}\sqrt{2m_0(V_0 - E)}$ against x and considering the area under the curve.

Fowler and Nordheim have applied this method of solution to the phenomenon investigated by Millikan and others of the extraction of electrons from cold metals in a vacuum by the application of very strong electric fields. The result they obtained, which is in agreement with experiment, is a triumph for wave mechanics, because the classical mechanics failed to elucidate the problem.

Another problem of this kind is that which forms the subject of this section. The actual problem is, of course, more complicated, the potential curve of Fig. 10 being replaced by that of Fig. 9, and the particle being bounded all round by a potential barrier. If the particle can move between these barriers on the average over a distance d, we can regard the probability of its escape at each approach to the barrier as being proportional to e^{-J}. The frequency of its approach to the barrier is $\frac{d}{v}$, where v is its velocity. Thus the average time that elapses before it escapes is of the order $T = \frac{d}{v}e^{J}$.

Now where V is zero all the energy is kinetic, and we have $\frac{1}{2}m_0v^2 = E$; thus

$$T \sim d\sqrt{\frac{m_0}{2E}}e^{J} \quad . \quad . \quad . \quad (3{\cdot}36)$$

If there be a large number, N, of particles with the same energy, E, we can write for the number, dN, which escape in time, dt,

$$dN = -\lambda N dt,$$

which leads to the radioactive law

$$N = N_0e^{-\lambda t},$$

with the average life

$$T = \frac{1}{\lambda}.$$

Thus

$$\lambda \sim e^{-J}.$$

One of the puzzling things about radioactive elements is the great differences exhibited in their rates of decay. Thus radium C′ has an average life of a minute fraction of a second, radium A of 4·4 minutes, and uranium of 6000×10^6 years.

Condon and Gurney, in their work referred to above, have plotted curves for $\frac{4\pi}{h}\sqrt{2m_0(V_0 - E)}$ against x, and

their results show that these great variations are exactly what the theory predicts. The empirical formula of Geiger and Nuttall for the constant λ is

$$\log \lambda = a + bE,$$

where b is a constant for all the radioactive families and a has a value peculiar to the family. E denotes the energy of the particle emitted.

Gamow has shown that the theoretical value of λ leads to this law.

CHAPTER IV

EXPERIMENTAL EVIDENCE OF THE WAVE-LIKE CHARACTER OF MATTER

THE preceding chapters have shown us that, if we adopt a certain equation and apply it to problems occurring in connection with the atom, we are able to calculate certain energy values in agreement with observation in atomic physics.

In a particular example it has been deduced that the energy levels of the hydrogen atom are given by

$$E_\tau = -\frac{2\pi^2 me^4}{h^2\tau^2}$$

in agreement with Bohr's theory.

The equation employed, as we have seen, is a wave equation, and we have introduced it as an addition to the familiar mechanics of a particle, and the study of the atom is on the one hand the study of a particle, and on the other the study of a wave. This is what the new mechanics suggests.

Experiment has left no doubt in our mind of the existence of the corpuscular aspect of matter. It has, in fact, until recent times, driven any other conception out of our minds.

Theoretical investigations invite us to contemplate the undulatory aspect of particles of matter, but we must remember that the wave-form of Schrödinger's equation may be fortuitous, and so we must beware of an immediate assumption of the existence of waves.

We are forced by the theory to make an appeal to

experiment, and by analogy with optics the path t
is clear.

The wave equation suggests the association of
length $\lambda = \frac{h}{mv}$ with the particle. This follov
(2·3), since

$$\lambda\nu = U = \frac{W}{G} = \frac{h\nu}{G}$$

$$\lambda = \frac{h}{G} = \frac{h}{mv}$$

In experiments upon electrons this wave-lengt
the same order as that of X-rays, and this s
experiments on an electron beam similar to thc
have been carried out on X-ray beams for the
mination of X-ray wave-lengths, e.g. diffract
crystals and ruled gratings.

The earliest experiments of this kind which e
the wave-like character of electrons are those of D

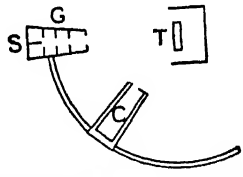


FIG. 11.—Diagram illustrating Davisson and Germer's ap

and Germer (*Nature*, 16th April, 1927 ; *Physical*.
vol. 30, p. 706). Earlier work by Davisson and Ku
(*Physical Review*, vol. 22, p. 242), carried out i
suggests that there was already evidence for th
nature of particles when the new theory appear
it is the opinion of Davisson that the results obta
that time cannot be regarded as properly exhibit
effect. The experiments of 1927 open a new cha
the history of experimental physics, and it is imp
to overstate their importance.

The electrons were emitted from a filament S, of heated tungsten ribbon, passed through slits to produce a narrow beam, and made to strike the target T normally. They were then collected in a receptacle C, which could move on an arc over such a range that beams could be received making angles between 20° and 90° with the incident beam. A sensitive galvanometer was connected to C and the deflections recorded were proportional to the number of electrons received. The potentials were arranged so that only those which had suffered a small loss of velocity at the target could enter the collector.

The bombarding current was of the order of a microampere, and of this about 10^{-4} was diffracted to the collector. A sensitive galvanometer could readily detect a current of this magnitude. The bombarding potentials ranged from 15 to 350 volts.

The target could be rotated about an axis perpendicular to its surface so that various crystal planes could be made to pass through the collector. The three crystal planes concerned in the experiments were the (111), (110) and (100) planes (see Fig. 15). The collector was moved to various positions in these planes and the galvanometer current was plotted against the angle between the primary and diffracted beam (Fig. 11). This angle was described as the "colatitude."

Fig. 12 shows the result obtained in the case of the (111) plane. As the voltages vary a peak appears at about 44 volts, and attains its greatest development at 54 volts in colatitude 50° (Fig. 12). If the voltages are increased the peak ultimately dies away, and there is hardly any trace of it at 68 volts.

If the crystal be rotated into the other two planes mentioned, small peaks are shown at 50° for 54 volts. These smaller maxima correspond to peaks, which reach their greatest development at other voltages and colatitudes.

Three different types of interference phenomena were detected: (*a*) scattering from a space lattice analogous

to the X-ray phenomenon of Laue and Bragg; (*b*) scattering from a "plane grating" arising from a single layer of atoms situated at the surface; and (*c*) scattering from adsorbed gas on the surface of the crystal which was responsible for certain sets of observations. It was found possible to infer the structure of the gas film from them.

The type of interference under (*b*) is not observed in X-rays, and arises from the fact that electrons are much more readily absorbed than X-rays. They result from a preponderance of first layer scattering at angles near grazing emergence.

The results of the observations under (*a*) and (*b*) show that there is a one-to-one correspondence with Laue-

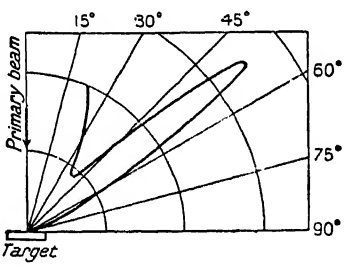


FIG. 12.—Diffraction of a beam of electrons for 54 volts bombarding potential ($\lambda = 1{\cdot}67$ Å).

Bragg beams which would issue from the same crystal if the electrons were X-rays, and that there is an analogy with optically diffracted beams from plane gratings.

Thus we have striking qualitative evidence in favour of the wave-like character of electrons.

The radii vectores to the curve measured from the target are proportional to the electrons collected in the corresponding directions.

The results turn out to be also in quantitative agreement. There was originally a systematic disagreement between the results obtained by the experimenters and

de Broglie's theoretical value (4·1), but the explanation which removes the difficulty may be mentioned because it adds to the analogy between particle waves and optical waves.

Eckart and Bethe [1] have drawn attention to the necessity of considering the refractive index of the waves.

In calculating Bragg's formula (see below)

$$2d \sin \theta = n\lambda \quad . \quad . \quad . \quad (4{\cdot}2)$$

for X-rays, we assume that the refractive index of the crystal is unity, and this is the case for X-rays. If the index is μ the formula becomes

$$2d\sqrt{\mu^2 - \cos^2 \theta} = n\lambda \quad . \quad . \quad (4{\cdot}3)$$

We have seen in Chap. II. (2·26) that the velocity of the wave is given by

$$U = \frac{h\nu}{\sqrt{2m(E - V)}} \quad . \quad . \quad (4{\cdot}4)$$

in a field where the potential energy is V, for example inside the crystal. Outside, where the field is zero,

$$U_0 = \frac{h\nu}{\sqrt{2mE}} \quad . \quad . \quad . \quad (4{\cdot}4\,1)$$

Thus the refractive index is

$$\mu = \frac{U_0}{U} = \sqrt{\frac{E - V}{E}} \quad . \quad . \quad (4{\cdot}5)$$

In the experiments of Davisson and Germer E was of such a value that μ could have differed from unity in some cases by about $\frac{1}{10}$. In the experiments to be described below, where the electrons moved much more rapidly, the value of E is greater and μ does not differ from unity by an appreciable amount.

In the experiment just described the diffraction was caused by a single crystal. A very remarkable series of

[1] *Naturwiss.*, 15, p. 787, 1927.

experiments has been carried out by G. P. Thomson in which the electron stream was directed upon a thin film composed of many small crystals.

The electrons were subjected to a fall of potential of the order of 25,000 volts. It is interesting to calculate the wave-length corresponding to this case, and we will make the calculation without entering into the refinement introduced by the relativity formula connecting mass and speed.

Let m denote the mass of the electron, and v the velocity acquired in the field. Let P denote the potential difference in absolute units.

Thus $$\tfrac{1}{2}mv^2 = e\mathrm{P} \quad . \quad . \quad . \quad (4{\cdot}6)$$

whence $$mv = \sqrt{2me\mathrm{P}} \quad . \quad . \quad (4{\cdot}7)$$

and $$\lambda = \frac{h}{\sqrt{2me\mathrm{P}}} \quad . \quad . \quad (4{\cdot}8)$$

If the appropriate values be substituted, the wave-length obtained is $\cdot 78 \times 10^{-9}$ cm., or $\cdot 078$ Ångströms. This is equal to the wave-length of very hard X-rays.

For voltages of the order of 100 such as those considered in the experiment described already, the wave-length is 1·22 Ångströms. While Davisson and Germer's experiment is analogous to the single crystal experiments of Laue and Bragg, Thomson's experiment is analogous to that associated with the names of Debye, Hull and Sherrer, in which X-rays were reflected at the surfaces of small crystals arranged at random. The resulting pattern on a photographic plate shows a definite symmetry if certain crystal directions predominate.

Thomson caused a beam of cathode rays, made approximately homogeneous, to pass normally through a very thin film. The rays were then received on a photographic plate and the pattern studied.

The apparatus is of a simple character, the rays being generated by an indication coil and caused to pass through a fine tube, then through the film to be intercepted by a plate after travelling a distance of about 32·5 cm.

ing the plate a symmetrical pattern about
t was observed similar to that observed

g this as evidence in favour of the presence
waves, we must be convinced that particles
te and produce the pattern and that the
al.
ivable that the cathode particles produce
these give the wave pattern, but this can
y generating a magnetic field between the
e. It was found that the pattern moved
der the influence of the field as if the beam
ie plate consisted of charged particles all
ie velocity. This differentiated the beam

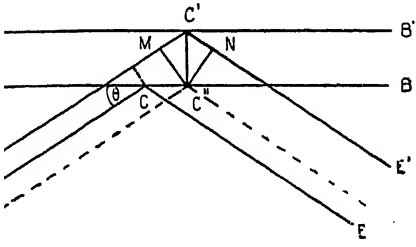


FIG. 13.

sisting of rays such as X-rays, and it was
r the homogeneity of the beam.
d to the other point, it need only be stated
ving the film only the central spot appeared
nd no pattern.
triking thing about this experiment is the
greement with our theory.
ate the quantitative agreement we must
g's theory of crystal reflection. According
y a crystal consists of atoms arranged in a
each atom forming a centre at which the
ected; from this it follows that we may
tal as made up of reflecting planes oriented

in certain directions. These planes pass through the atoms, and on account of the symmetry in the crystal structure, sets of parallel planes corresponding to the different layers of atoms occur ; for example, if we think of the atoms arranged at the corners of a cube reflecting planes will correspond to the sides and diagonal planes.

Let AB and A'B' denote two consecutive parallel planes, and let two rays be incident at C and C' and be reflected along CE and C'E'. If these rays are in the same phase they will reinforce one another and produce an intense line on a photographic plate placed to receive them.

The difference in path between them is the same as that between the dotted path and the second ray. Let C''M and C''N be drawn perpendicularly to C'D' and C'E' respectively. The path difference is MC' + C'N. If θ denote the angle between the rays and the planes, and d the distance between the planes, this difference is equal to $2d \sin \theta$.

Thus the condition for reinforcement is

$$2d \sin \theta = n\lambda, \ (n = 1, 2, 3, \ldots) \quad . \quad (4{\cdot}9)$$

The value of d differs according to the planes considered.

In the case of a cubic arrangement it is easy to express d in terms of the side of the cube, say a. Suppose that three directions parallel to the sides of the cubes are taken as axes, then the equation of a plane referred to these may be written : $lx + my + nz = p$, where (l, m, n) denote the direction cosines of the normal to the plane and p is the perpendicular from the origin on to the plane.

Suppose that we think of a plane through the origin as one of our reflecting planes with an adjacent parallel plane at distance d.

The equation for the adjacent plane is $lx + my + nz = d$, and we could distinguish our plane by the set of direction cosines $(l\ m\ n)$. Instead of this we take three integers $(h_1\ h_2\ h_3)$ and distinguish the plane by speaking

of it as the '$h_1h_2h_3$ plane' where the equation of the plane is thrown into the form

$$h_1x + h_2y + h_3z = a.$$

We have, of course,

$$\frac{h_1}{l} = \frac{h_2}{m} = \frac{h_3}{n} = \frac{a}{d},$$

and since $l^2 + m^2 + n^2 = 1$ we have

$$d = \frac{a}{\sqrt{h_1^2 + h_2^2 + h_3^2}} \quad . \quad . \quad (4{\cdot}10)$$

which gives d in terms of a from the plane indices $(h_1h_2h_3)$.

Suppose that a thin film containing small crystals in all possible orientations at random receives a beam of X-rays, then by (4·2) we have reinforcement if

$$2d \sin \theta = n\lambda,$$

where d is the separation of the reflecting planes concerned and θ is the glancing angle on a small crystal.

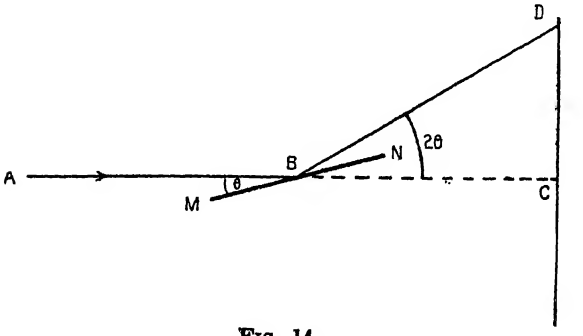


FIG. 14.

If the reflected beam is received at D on a photographic plate and the angle is small, we have

$$CD = 2L\theta,$$

where C is the point of incidence of the primary on the plate in the absence of the crystal MN and L is the distance BC, i.e. the distance from film to plate.

Thus from (4·9) we have an intense spot at D if

$$CD = \frac{n\lambda L}{d} \quad . \quad . \quad . \quad (4\cdot11)$$

If we think of B as the apex of a cone with its base to the right of the diagram, and if crystals lie so that the planes MN are tangent to the cone at the apex, a small pencil of rays along AB will fall at the same angle on all the crystal planes giving rise to a circle on the plate of radius CD. Since we assume that the crystals are arranged at random we have the equivalent of this reflecting cone.

For the same value of D we shall get a series of circles corresponding to the values of n. Other circles will be obtained for different planes MN, i.e. with different values of d, and it is possible that reflecting planes will occur in the crystal not much inclined to MN at which reflection of this nature can occur, the different values of θ required occurring in the beam AB, which cannot be quite parallel in practice.

If the arrangement is not at random the circles may be incomplete, and at any rate will show different intensities at different parts of the ring.

In order to put the theory to the test by verifying the equation (4·1), we have to measure the diameters of the rings formed, determine the planes responsible for the diffraction and deduce λ from (4·11). The experimental result can then be compared with the theoretical value.

The method used by Thomson was essentially this, but he made the test by calculating the constant a from the mechanical wave-length of the cathode rays and comparing this with the value determined in experiments on crystal structure.

It is very interesting and instructive to refer to the original paper on this work for the photographs of the patterns obtained. One of these is for a gold film and

the rings are remarkably clear and uniform in intensity. The diameters of these rings are approximately in the ratio 1 : 1·18 : 1·62 : 1·96. We have to determine if the structure of gold crystals is such that reflecting planes occur which would produce rings with diameters in these ratios. Gold crystals are built up as a face-centred cubic lattice and for the sake of simplicity we will take three planes with indices (2, 0, 0), (2, 2, 0) and (1, 1, 3) and find the ratios of the diameters to be expected. If we refer to (4·11) and consider the first order ring ($n = 1$) the ratios are inversely as the corresponding values of d, and that according to (4·10) is in the ratio of the corresponding values of $\sqrt{h_1^2 + h_2^2 + h_3^2}$ so that in this case the ratios are

$$\begin{aligned} & \sqrt{4} : \sqrt{8} : \sqrt{11} \\ = \quad & 2 : 2{\cdot}83 : 3{\cdot}32 \\ = \quad & 1 : 1{\cdot}41 : 1{\cdot}66 \end{aligned}$$

We may write the ratios of the last three observed rings, given above, as

$$1 : 1{\cdot}39 : 1{\cdot}66$$

so that we may claim that these observed rings are accounted for. By examination of other planes it was found possible to account for the first ring.

A film of aluminium gave equally convincing evidence.

Another striking quantitative result can be obtained by observing that if D denote the diameter of a ring on the pattern $\frac{D}{\lambda}$ should be constant for the same order n (4·11).

Thus by varying λ and observing the value of D we can test this result.

There is no point in making a relativity correction here so that we may use the value given above, viz. :

$$\lambda = \frac{h}{\sqrt{2mPe}},$$

which shows that the theory anticipates that $D\sqrt{P}$ is a constant. Thus by measuring P, the potential applied to the cathode rays, by varying it and measuring the diameters it is easy to test the result.

A very good agreement was found, as will be seen by reference to the original paper where the results are given for aluminium and gold. The agreement here is strong evidence in favour of the wave-like behaviour of the electrons.

With regard to the determination of the value of a by this experiment, the X-ray value for gold is $4{\cdot}065 \times 10^{-8}$ cm., and the value obtained is within about 1 per cent. of this.

Experiments of the same kind as the foregoing have been carried out by many workers with the result that the law of wave-length has now been firmly established and has become known as the de Broglie law.

The experiments of M. Ponte, described in the *Comptes Rendus*, **188**, p. 244, in 1929, though based on the same principle as that of the Debye-Sherrer experiment, must be mentioned because of the great improvement in a point of technique.

The method of G. P. Thomson requires the deposition of a thin crystallised film, and requires a delicate and difficult technique.

In Ponte's experiments crystalline powders of zinc, magnesium and cadmium oxides were used. A narrow beam of electrons from a hot cathode was directed on to a thin metallic wire or on the edge of a diaphragm on which the oxide had been formed. The beam was thus not required to pass through the film, and the very delicate deposition of a thin film was not necessary.

In this way by using tensions varying between 7·6 and 17·25 kilovolts as many as 20 rings were determined and crystal dimensions calculated. The results obtained agree within 1 per cent. with those of Bragg.

The Application of Electronic Diffraction to the Study of Surface Phenomena

The interference phenomena occurring in the reflection of slow electrons open new possibilities for the examination of surface structure. They have two advantages over X-rays in this respect. They penetrate less deeply and the interaction between the electrons and the atoms of the experimental body is more intimate than that occurring in the case of X-rays. In the case of electrons only the first ten or twenty planes are involved, while in the case of X-rays a thousand and more come into consideration.

The electronic interaction with the atoms shows itself by the occurrence of a refractive index for slow electrons.

In Davisson and Germer's experiment there occurred certain characteristics in the diffraction pattern which were attributed to several layers of gas atoms at the crystal surface.

The investigation of these layers has been carried out in later work by Germer and Rupp.[1]

The effect of allowing gas to come into contact with a Ni surface in Davisson and Germer's experiment was to diminish the intensity of the diffraction maxima, and this was due to layers of gas molecules covering the surface.

In addition to this intensity change, two other types of diffraction pattern were observed. These appeared after heating a crystal and allowing it to cool. They were of short duration and were noticeable a short time after the heating process.

Germer has plotted on a graph a series of values of the wave-length against $\sin \theta$. (θ is the angle of diffraction.) For a diffraction pattern the relation should be linear. He has found that from the nickel itself lines occur corresponding to the orders $n = 1, 2, 3$, while the diffraction

[1] Germer (*Z. f. Ph.*, 54, p. 408, 1929). Rupp (*A. d. Ph.*, 5, p. 453, 1930).

pattern of one of the two types above-mentioned corresponds to the order $\frac{1}{2}$. We have in this case to do with a plane grating with an element of double the magnitude of that responsible for the other orders.

Germer has concluded that the gas atoms form a single layer arranged like the atoms of the crystal, but with a double separation, and situated at a depth of 3 Å below the first layer of Ni atoms.

The arrangement is illustrated in Fig. 15.

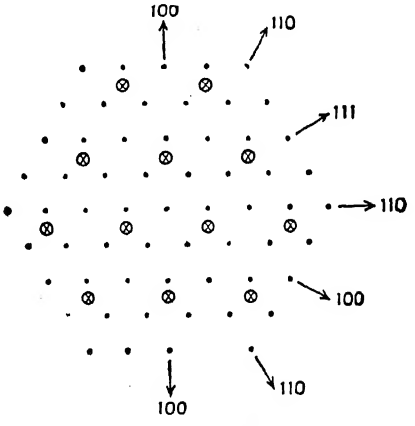


FIG. 15.—Arrangement of absorbed gas atoms at the surface of a Ni crystal. · Ni atoms. ⊗ Gas atoms.

In Rupp's experiments a stream of electrons was allowed to fall on a nickel surface and the potential driving them varied. The electrons fell into a metal receiver after reflection, and the intensities were measured corresponding to different wave-lengths, i.e. to different values of the potential V.

Curves were plotted showing the variation of intensity with $\sqrt{V}$ as in Figs. 16 and 17.

The angle of incidence was maintained constant, and the receiver was placed so that the ray entering it left

the crystal at an angle equal to the incident angle. Fig. 16 shows that for voltages of 67, 132, 215, i.e. for certain values of λ, the intensity attains maximum values. The corresponding orders are 3, 4 and 5.

After obtaining the pure Ni curve hydrogen was allowed

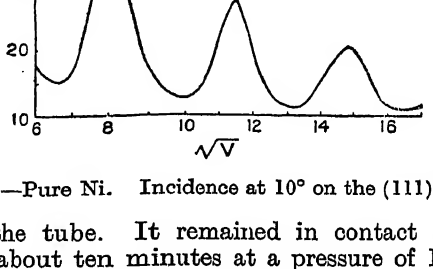


FIG. 16.—Pure Ni. Incidence at 10° on the (111) face.

to enter the tube. It remained in contact with the metal for about ten minutes at a pressure of 10^{-4} mm. Hg and was then pumped off.

About half an hour after the entrance of hydrogen the electron reflection was measured, beginning with low

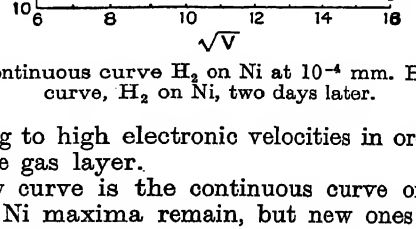


FIG. 17.—Continuous curve H_2 on Ni at 10^{-4} mm. Hg. Dotted curve, H_2 on Ni, two days later.

and passing to high electronic velocities in order not to disturb the gas layer.

The new curve is the continuous curve of Fig. 17. The main Ni maxima remain, but new ones have put in their appearance as subsidiary maxima. These new maxima are situated at points corresponding to voltages of 41, 96, 170, and the orders are $2\frac{1}{2}$, $3\frac{1}{2}$ and $4\frac{1}{2}$.

The dotted curve, taken two days later, shows a change in the intensities in the Ni maxima, and it is thought that this is due to the fact that the gas layer has become less regular, but the experiment gives no detailed information on this point, and Rupp describes the process as " surface loosening."

Various substances were investigated in the same way, argon on nickel, hydrogen on copper, etc.

It was found that there was no interaction between argon and nickel, but results were obtained in the case of hydrogen on other metals which resembled those of hydrogen on nickel. Difficulties occurred in some cases, and this was especially the case with diamond. The phenomenon did not appear to fit in with a diffraction equation.

This equation was obtained from (4·3) by substituting the value for μ which we obtained in (4·5),

i.e. $$\mu = \sqrt{1 + \frac{E_0}{P}},$$

where E_0 is the so-called grating potential $(-V)$ measured in volts, and $P\ (= E)$ is the voltage drop of the electron stream. This equation and (4·3) lead to

$$E_0 = \left(\frac{n}{2d}\right)^2 \frac{h^2}{2em} - P \sin^2 \theta \qquad . \quad (4·12)$$

On substituting the values for h, e and m this equation becomes

$$E_0 = \frac{150}{4d^2} n^2 - P \sin^2 \theta \qquad . \quad . \quad (4·13)$$

where d is measured in Ångströms.

The procedure for the determination of E_0 is to try small integral values of n, taking the corresponding values of P from the maxima of a curve such as that of Fig. 16. When the correct values of n have been found in this way, the value of E_0 is the same for the various maxima. d is obtained from X-ray analysis.

In some cases it is necessary to take half odd integer values for n in order to arrive at a definite value for E_0. This is the case in the example given above.

These half integer values are striking, and empirically there is no doubt about their occurrence, which may be due to the presence of foreign gas atoms, as Germer has explained. But it is strange that the grating constant d occurs again so simply in the formula.

We must also consider the possibility of something corresponding to phase change at the surface, as in the similar optical phenomenon. In this case only the half numbers occur. There is evidently much to be explained before the treatment can be considered satisfactory. The value of E_0 is not the same as the photoelectric threshold potential, and the reason may be that the work required to remove an electron across the surface differs from the potential drop on account of the kinetic energy it has inside the substance. Some of the values obtained by Rupp are 16·5 volts for Ni, 14·5 volts for Fe, and 12·8 volts for Cu.

Diffraction of Electrons by a Ruled Grating

The fact that electrons and X-rays are diffracted in the same way by crystals suggests that there may be a similarity in the case of diffraction by a ruled grating.

The possibility of carrying out experiments of this kind has been examined by Rupp [1] and Worsnop.[2] The following is an account of Rupp's experiment, which again verified the existence of electronic wave properties, the wave-length being given by de Broglie's formula.

Fig. 18 illustrates the apparatus diagrammatically. A tungsten wire covered with barium oxide was placed at K, and between this wire and a second wire G there was a tension of about 12 volts. From the second wire to the shutter S_1 the chief potential drop was maintained, and it could be varied.

[1] *Zeit. f. Phys.*, 52, p. 8.
[2] *Proc. Phys. Soc.*, xxxvi, p. 284. *Nature*, Feb. 2, 1929.

In this way the film was protected from direct light from the hot wire.

S_1, S_2 and S_3 denote three shutters, and the magnetic coil MM focussed electrons from S_1 on the film at F.

Beyond S_3 the electrons struck the grating G, which could be rotated and set at an angle to the axis by means of a micrometer screw.

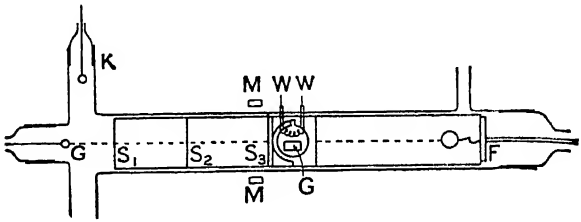


FIG. 18.

The coil WW represents a tungsten spiral for bombardment of the grating.

The distance S_1S_2 was 32 cm., and from the grating to the film 38·5 cm. Let an incident ray S_1G strike the grating at the grazing angle θ, and let GF_1 denote the

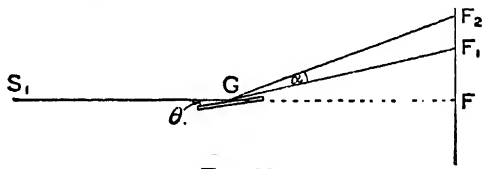


FIG. 19.

reflected ray. Let GF_2 denote a diffracted ray, making an angle α with GF_1.

If we apply the diffraction grating formula to this case, bearing in mind that the angles are small, we find

$$\tfrac{1}{2}d\alpha(\alpha + 2\theta) = n\lambda,$$

where d is the grating constant.

In the grating employed by Rupp $d = 7{\cdot}70 \times 10^{-4}$ cm. measured by means of the mercury green line.

The experimental results may be examined in two ways. The first is a relative method, and consists in maintaining θ constant, while the electronic velocity is varied. From the experiment the value of θ may be deduced for each velocity, i.e. each λ, and the results should agree in all the cases. The experiment gives α and $\lambda = \frac{h}{mv}$, so that θ may be calculated from the above formula.

The second is an absolute method. It consists in observing the different orders in the diffraction pattern and the corresponding angles of diffraction. From these observations both θ and λ can be calculated.

The reason for this procedure is that the grazing angle is not able to be measured with the necessary accuracy. According to Rupp, the experimental results verified the theory to an accuracy of about 2 per cent.

Experiments on Neutral Particles

In searching for another opportunity to test the de Broglie wave-length relation, it is natural to examine the possibility of making use of gaseous molecules. In this case the mass of the particle is large, but the velocity is comparatively small. The value of the wave-length in the case of hydrogen at room temperature is approximately 10^{-8} cm. for the molecules moving with the most probable velocity.

T. H. Johnson (*Jour. Franklin Inst.*, **206**, p. 301, 1928) investigated the reflection of atoms by crystals, using a plate smoked with molybdenum trioxide as a detector. This substance becomes blackened where hydrogen atoms strike it. The results indicate that the atoms have a wave-like character.

Ellett, Olson and Zahl (*Phys. Rev.*, **34**, p. 493, 1929) have made experiments with mercury, cadmium and

arsenic beams, which they caused to fall on a rock-salt surface.

They conclude from their results that the reflection is of the type associated with the name of Bragg in the case of X-rays. They have shown that a beam reflected by a crystal acquires a property by virtue of which it is entirely reflected by a second crystal only when the second angle of incidence is equal to the first. The beam is in some way prepared by the first crystal for this particular type of reflection by the second.

They have also shown that the reflected beam is made up of atoms all with very nearly the same velocity, which varies with the angle of reflection, and that the intensity of the reflected beam also varies with the angle of reflection.

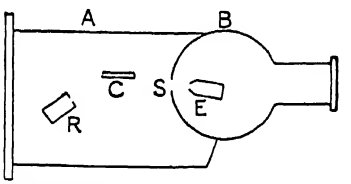


FIG. 20.—Diagram to illustrate the apparatus of Stern and Estermann.

In these experiments the reflected atoms are detected by receiving them on a glass surface cooled by liquid air.

Stern, Knauer and Estermann have conducted a series of experiments on the diffraction of gas molecules, and we shall give a short account of experiments described in one of their recent papers (Stern and Estermann, *Zeit. für Physik*, **61**, p. 95, 1930). The apparatus consisted of two main chambers A and B, separated by a partition carrying a slit S. The apparatus is represented diagrammatically in Fig. 20. E is an ejector from which the molecules are shot through the slit on to the crystal C, and R is a receiver which can be adjusted to receive the molecules as they come from the crystal.

The pressure in chamber A was of the order of 10^{-5} mm. Hg, and in B of the order 10^{-4} mm. Hg. These pressures were recorded on a Macleod gauge.

The receiver was connected to a manometer which measured the changes in pressure due to the incidence of the beam of molecules. This manometer is a very important part of the apparatus, and it must be sufficiently sensitive to determine changes of pressure of the order 10^{-8} mm. Hg. The pressure prevailing is of the order 10^{-5} mm. Hg.

The experimenters finally decided upon the Piriani hot-wire manometer as being the most suitable instrument, and they describe some improvements for the increase in the sensitivity of the instrument.

The gases used were hydrogen and helium, and the velocity of ejection from E was controlled by means of the temperature.

The distribution of velocities is known from the Maxwellian law, and this can be translated into a distribution among wave-lengths by applying de Broglie's relation.

It can be shown that the wave-length of greatest intensity is given by

$$\lambda_m = 19{\cdot}47 \times 10^{-8} \times \frac{1}{\sqrt{\mathrm{T}m}} \text{ cm.},$$

where T denotes the absolute temperature and m the mass of a molecule. It seems impossible to think of the process as one of diffraction at a space grating, since the molecules cannot penetrate the crystal. We must regard the phenomenon as one of diffraction at a surface if it is diffraction at all.

Let the z-axis lie perpendicularly to the crystal surface, and let the x- and y-axes lie along the two principal axes of the grating. Consider a beam of parallel rays falling upon a grating, one element of which is represented in Fig. 21 by ABCD.

Imagine a wave-front drawn at right angles to the bundle of rays aA, bB, cC and dD, and a second wave-front at right angles to Aa', Bb', Cc' and Dd'.

The paths of aAa' and bBb' between these wave-fronts differ by $(\cos\alpha - \cos\alpha_0)$ AB, where α_0 and α are the angles made by the incident and diffracted rays with AB.

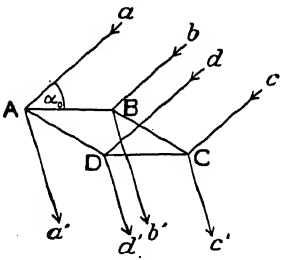


FIG. 21.

Similarly, the path difference for the rays cCc' and bBb' is $(\cos\beta - \cos\beta_0)$BC, where β_0 and β are the angles made with BC.

In any direction in which these path differences amount to a whole number of wave-lengths we shall find reinforcement of vibrations, and thus the diffraction pattern is determined by the equations

$$\cos\alpha - \cos\alpha_0 = h_1\frac{\lambda}{d}$$

$$\cos\beta - \cos\beta_0 = h_2\frac{\lambda}{d},$$

where h_1 and h_2 are integers (positive or negative), λ is the wave-length of the incident waves, and d the grating element, supposed the same in both directions, i.e.

$$\text{AB} = \text{BC} = d.$$

The orders of the diffracted rays are described by means of the integers, e.g. (h_1, h_2).

In the experiments of Stern and Estermann, the ray was generally incident in the xz-plane, so that β_0 had the value 90°.

If the values of h_1 and h_2 are both zero, $\alpha = \alpha_0$ and $\beta = 90°$. This corresponds to the reflected ray. In the experiment the diffracted rays of order (0 ± 1) were observed as well.

The diagram shows the result for the case of rays formed of helium molecules incident upon a crystal of

lithium fluoride at an angle of $11\frac{1}{2}°$ and at the temperature of 22° C.

The process consisted in directing rays at a constant angle of incidence upon the crystal and setting the receiver at an equal angle to the normal to the crystal surface.

The abscissa zero of the diagram corresponds to the case where the receiver lies in the plane of incidence.

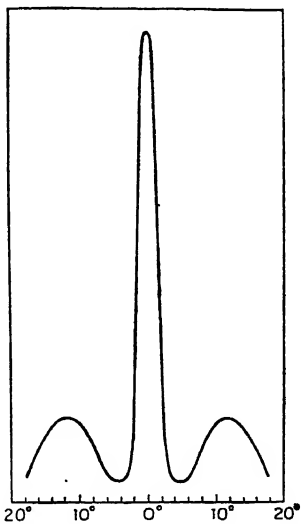


FIG. 22.—Diffraction of He by LiF.

By rotation of the receiver about the normal into different azimuths variations of pressure were detected. The abscissæ measure the azimuths and the ordinates are proportional to the manometer readings, i.e. to the ray intensities in the different directions.

Two subsidiary maxima occur on either side of the zero in azimuths of approximately 12°. The arrows

show the positions of the calculated maxima for the orders (0, $\pm$ 1), the angles are at $11\frac{3}{4}°$. The agreement is very good.

This curve is typical of many obtained, and recorded in the original paper.

CHAPTER V

THE PHYSICAL SIGNIFICANCE OF THE WAVE EQUATION AND THE WAVE FUNCTION

The Density and Probability Interpretation of ψ

IT is of the greatest interest and importance to inquire whether the wave function has a physical significance and to attempt to discover the true place of the wave equation in physics. It is not in the spirit of mathematical physics to introduce an isolated function and a differential equation into the theory and to leave it there without inquiring about its relation to the accepted doctrine.

We have developed a theory which leads us by analogy with optical theories to the wave equation and introduces a quantity, ψ, but, after all, we have not actually built the equation into the classical theory, and ψ has not been associated with any physical quantity.

Although we have always to begin with some fundamental assumptions we try to reduce the number to a minimum.

This has been illustrated in this book in the case of mechanics in our study of the Hamilton principle or in the example of the Hamilton-Jacobi equation. We may look upon Hamilton's principle, for example, as the fundamental assumption and from this deduce the laws of classical mechanics. Such principles play an important part in all branches of physics, and it does not surprise us that the wave equation can be derived from a principle of this kind. We may have to be content with this and look upon the equation as an isolated principle

with no connection with anything familiar to us, but we shall not do so without a struggle.

We propose in this chapter to review the work of embodying ψ and its equation into our physical doctrine.

First we consider Schrödinger's interpretation of ψ, which he adopts by analogy with the classical theory of radiation from a system of electrically-charged particles. In this theory the radiation is determined by the electric moment of the system of particles.

This quantity is measured by reference to three co-ordinate axes, and if the charges have x-co-ordinates x_1, x_2, x_3, etc., and magnitudes e_1, e_2, e_3, etc., the x-component of the moment is given by

$$p_1 = e_1x_1 + e_2x_2 + e_3x_3 + \text{etc.},$$

and there are similar y- and z-components.

If the charges oscillate, the co-ordinates will be given as some function of the time, and p_1 can be represented as a Fourier's series, which may be written thus:

$$p_1 = A_0 + A_1 \cos(2\pi\nu t + \delta_1) + A_2 \cos(4\pi\nu t + \delta_2) + \text{etc.} \quad (5{\cdot}1)$$

the general term being $A_n \cos(2\pi n\nu t + \delta_n)$, ν being the fundamental frequency of the system.

It is more convenient to express the series in another way:

$A_n \cos(2\pi n\nu t + \delta_n) = \frac{1}{2}(A_n e^{i\delta_n} e^{2\pi i n\nu t} + A_n e^{-i\delta_n} e^{-2\pi i n\nu t})$ where $i = \sqrt{-1}$.

A_n is a real quantity, and the only difference between $A_n e^{i\delta}$ and $A_n e^{-i\delta_n}$ is in the change of sign before the imaginary root. Quantities with this relation are said to be conjugates. We shall write C_n for the first and C_{-n} for the second, so that the general term may be written

$$C_n e^{2\pi i n\nu t} + C_{-n} e^{-2\pi i n\nu t}$$

and the value of p_1 is given by

$$p_1 = \sum_{-\infty}^{\infty} C_n e^{2\pi i n\nu t} \quad (5{\cdot}2)$$

the summation being extended over all positive and negative integer values of n.

Thus the moment may be considered as equivalent to the superposition of harmonic oscillations of frequencies ν, 2ν, etc., the amplitudes of these oscillations being measured by A_1, A_2, etc., and the energies being proportional to A_1^2, A_2^2, etc.

If we use the expression (5·2) instead of (5·1) from which the A's have disappeared, we shall express the energies or intensities of the corresponding component oscillations by C_1C_{-1}, C_2C_{-2}, etc., since we have $A_1^2 = 4C_1C_{-1}$ and in general $A_n^2 = 4C_nC_{-n}$, so that we measure the intensities conveniently by the product C_nC_{-n} to which they are proportional.

Since C_n and C_{-n} are complex conjugate quantities we shall adopt the usual notation : $C_nC_{-n} = |C_n|^2$.

The classical theory of radiation made the assumption that such a system would give rise to radiations of the frequencies ν, 2ν, etc., occurring in the series and that the intensities of the corresponding waves would be proportional to $|C_1|^2$, $|C_2|^2$, etc.

Bohr regarded the principle as essentially a quantum principle, and showed that from the quantum principle one derived the classical principle when the quantum numbers, n, were large.

This principle of Bohr's, known as the correspondence principle, is evidently of fundamental significance, and Schrödinger recognises this by adopting an expression corresponding to (5·2) to represent the state of a system in relation to the radiation it will emit.

The series he adopts is not a Fourier's series, though it is of the same type. Instead of using sine and cosine terms he makes use of characteristic functions and the associated characteristic values. Let these be ψ_1, ψ_2, etc., and E_1, E_2, etc.

Then it is assumed that the expression

$$\psi = c_1\psi_1 e^{\frac{2\pi i}{h}E_1 t} + c_2\psi_2 e^{\frac{2\pi i}{h}E_2 t} + \text{etc.}$$
$$= \Sigma c_n\psi_n e^{\frac{2\pi i}{h}E_n t}$$

where the c's are constants, describes the radiation state.

The conjugate of ψ is written $\bar{\psi}$ and is evidently

$$\bar{\psi} = \Sigma c_n \bar{\psi}_n e^{-\frac{2\pi i}{h} \mathrm{E}_n t}.$$

Schrödinger makes the assumption that $\psi\bar{\psi}$ represents the density of charge and implies that the system consists of a continuous distribution of charge.

This may seem very arbitrary, but it is in keeping with his theory, which is one of continuities.

This view then replaces the discrete charges with a continuous density, and the electric moment is no longer to be written

$$p_1 = \Sigma e_n x_n$$

but is replaced by

$$p_1 = \int \rho x dv,$$

or by

$$p_1 = \int \psi \bar{\psi} x dv,$$

because of the assumption concerning $\psi\bar{\psi}$.

The product $\psi\bar{\psi}$ contains as a typical term

$$c_m c_n (\psi_m \bar{\psi}_n e^{2\pi i \nu_{mn} t} + \psi_n \bar{\psi}_m e^{-2\pi i \nu_{mn} t})$$

where $\nu_{mn} = \dfrac{\mathrm{E}_m - \mathrm{E}_n}{h}$, and thus denotes the frequency of radiation corresponding to a transition between two states according to the view of the old quantum theory.

Thus a typical term in the expression for p_1 is

$$c_m c_n e^{2\pi i \nu_{mn} t} \int \psi_m \bar{\psi}_n x dv + c_m c_n e^{-2\pi i \nu_{mn} t} \int \psi_n \bar{\psi}_m x dv.$$

The integrals are denoted by x_{mn} and x_{nm} respectively, and we have from

$$x_{mn} = \int \psi_m \bar{\psi}_n x dv \text{ the result } \bar{x}_{mn} = \int \bar{\psi}_m \psi_n x dv = x_{nm}.$$

Thus $p_1 = \Sigma c_m c_n (x_{mn} e^{2\pi i \nu_{mn} t} + x_{nm} e^{-2\pi i \nu_{mn} t})$

$$= \Sigma c_m c_n (x_{mn} e^{2\pi i \nu_{mn} t} + \bar{x}_{mn} e^{-2\pi i \nu_{mn} t}) \quad . \quad (5{\cdot}3)$$

In this we see an exact correspondence with the classical equation (5·2) where the coefficient associated with the exponential containing the negative power is the conjugate of that associated with the positive.

We thus replace the classical principle by (5·3) and suppose that the system gives rise to radiation of frequencies ν_{mn} and corresponding intensities proportional to $|x_{mn}|^2$.

These quantities, x_{mn}, are the components of the matrices introduced by Heisenberg in his discontinuous theory, and the relations above show how these components are calculated from the characteristic functions.

We obtain from this a remarkable generalisation of the classical and old quantum principles and must seek for an experimental verification.

An application has been made to the components of the lines H_α, H_β, H_γ and H_δ appearing in the Stark effect, and the agreement is better than that obtained by Bohr and Kramers who based their calculations on the older principle.

There is another interpretation of ψ which is different from that we have just been considering; this is the interpretation of Bohr and Heisenberg, and rests on a theory very different from that of Schrödinger.

It is assumed that problems on electrons are essentially of a statistical character, and the product $\psi\bar{\psi}$ measures a probability factor; for instance, if we think of an electron in space mapped out by the usual three co-ordinate axes, the probability that the electron lies in the element $dx\,dy\,dz$ is $\psi\bar{\psi}\,dx\,dy\,dz$.

Since the electron is assumed to lie in a certain region of space, it follows that the integral of this expression taken over the space is unity. This means that some appropriate factor must be applied to ψ to make the integral $\int \psi\bar{\psi}dxdydz = 1$. This does not affect the fact that ψ is a solution of the wave equation, and ψ, when determined to satisfy this integral relation as well is said to be normalized.

The interpretation of ψ as a probability function is based upon an assumption which is fundamentally different from that which underlies the view we develop in this book.

Bohr and Heisenberg state that there is a fundamental indefiniteness and a fundamental discontinuity associated with all quantities which we seek to observe in experimental physics. This is illustrated by an example which was given by Heisenberg, and although it is of a very special character there is nothing in the result which is a consequence of the special circumstances existing. The result is quite general.

An experiment is imagined for the determination of the position of an electron, and it is shown that although we may conceivably locate the electron to any order of accuracy we please we cannot at the same time observe both its position and momentum without incurring an error. The more accurately we locate the electron the more inaccurately do we determine its momentum, and conversely, and if q_1 denote the error in determining the position and p_1 the error in the determination of the momentum, then the product $p_1 q_1$ is of the order h. This, according to Bohr and Heisenberg, is the real significance of Planck's constant.

The special case considered by Heisenberg is that in which the electron is located by on optical method. We know that the error in determining the position of a particle in a microscope is of the order of the wave-length of the light used, and it is necessary in the attempt to observe very small objects to use light of very short wave-length.

Two points cannot be resolved or recognized as distinct by means of a lens if they are closer together than $\frac{\lambda}{A}$, where λ is the wave-length and A the aperture or the angle subtended by the lens at either of the two points.

The error in location is then of the order $\frac{\lambda}{A}$.

Now let it be imagined that radiation is used to locate an electron in this way. The experiment and the apparatus are, of course, purely imaginary, but we will speak of the location of the electron by the gamma-ray microscope, since gamma rays have very short wavelengths.

Then our quantity q_1, the error in the determination of the position of the electron, is of the order $\frac{\lambda}{A}$ and we write

$$q_1 \sim \frac{\lambda}{A}.$$

Now the rays which fall upon the electron give it momentum; we have in fact an example of the Compton effect, and Compton's calculation in which he attributes

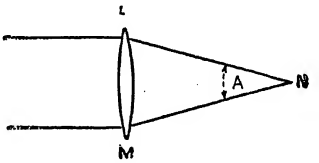


FIG. 23.

to the radiation a momentum $\frac{h\nu}{c}$ is verified experimentally.

Part of this momentum is taken up by the electron so that in the act of locating it its momentum is changed.

Now the direction of the radiation lies within the angle A, and we do not know the direction any more accurately than this. At the worst we do not distinguish between the direction LN and MN along which the quanta may come.

Thus we do not distinguish between the momentum p along, say, AB and p along CB, and consequently

make a mistake, which at its maximum is equal to AC and is thus of magnitude pA since A is a small angle

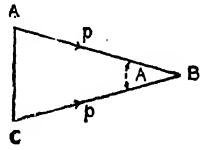


FIG. 24.

Thus the error in p is of the order pA and we write

$$p_1 \sim p\mathrm{A}$$

or

$$p_1 \sim \frac{h\nu}{c}\mathrm{A},$$

ν being the frequency of the radiation and equal to $\frac{c}{\lambda}$ where c is the velocity of the radiation. Thus

$$p_1 q_1 \sim h.$$

With this inherent uncertainty in our problems it is necessary to take a statistical view so that the introduction of a probability function is essential; this function, as we have seen, is $\psi\bar{\psi}$.

THE QUESTION OF THE EXISTENCE OF WAVES

Experiments upon light waves, upon X-rays and gamma rays have made us believe in the actual existence of electromagnetic waves. If we think of them as ether waves we imagine that there are actual ether vibrations associated with them. The difficulty here has been to associate a corpuscular character with the obvious wave character. We have, however, now definitely recognized that waves of this kind present themselves in the double guise of extended wave and of corpuscle. The work of de Broglie has taught us that this duality is not confined to light and X-rays. We must recognize it everywhere in physics. The particles of mechanics to

which we readily attribute the corpuscular character have also a wave character. The difficulty here is of the opposite kind to that in the former case. The wave aspect has escaped us until the last few years. But the question now is whether these waves exist in the same sense that waves of light exist. The question cannot be answered definitely, but various views have been expressed.

Schrödinger advanced the idea that the electron is a group of waves, which he called a 'wave packet.' But it is difficult to understand how such a structure preserves its identity. Why is it not dispersed when it is diffracted by a crystal?

Another view is that the particle is a singularity in a wave, but although this idea can be developed in the case of uniform motion, it is impossible to generalise it.

This view was at one time put forward by de Broglie, who also made another suggestion, which was known as the theory of the pilot wave. The particle and wave were associated so that the latter was guided by the former. The wave acted as a pilot for the particle. But there are objections to this point of view also.

The point of view which appears at present most satisfactory is that of Bohr and Heisenberg. According to it the wave has no physical significance, and the wave equation is an equation which gives ψ, the probability function we have mentioned in connection with the uncertainty relations.

The patterns observed in the experiments we have described in an earlier chapter are a representation of probabilities, the electrons falling according to definite probability laws into the places where they are seen in the patterns.

The fact that the patterns are like the diffraction patterns familiar in association with X-rays is due to the fact that the probability function satisfies an equation which has a wave form.

Perhaps we ought to take the probability view of the X-ray patterns also, the particles in this case being

photons. But the photon occupies a rather special position for its wave equation indicates that it is a particle of negligibly small mass.

We have to admit that in both matter and radiation there exists a duality of wave and particle; the wave character is more obvious the lighter the particle. It is difficult to think of matter or radiation as being both wave and particle, and any attempt to do so springs from a wrong attitude to the problem. We are confronted here with a problem of a very fundamental character, and the nature of this duality is as yet by no means understood.

Much of the difficulty appears to lie in our predilection for continuity and for space-time description of phenomena. If we persist in describing phenomena according to the methods of classical physics by means of space and time, then we must give up our ideas of continuity. This follows from the discovery of the uncertainty relations.

If we wish to retain the conception of continuity in physics, we must give up space-time description. We cannot have both.

So fundamental a departure from classical ideas requires a new mathematical apparatus and a new outlook. We must not expect to be able easily to picture by means of models the fundamental things of nature. We must look upon the models as little more than aids to memory or summaries of what we know about the particular phenomenon we are describing.

Perhaps the most profound of the methods of approach to modern problems is that of Dirac who, in his *Quantum Mechanics*, builds up a logical theory for the description of phenomena. Certain functions are introduced for this description, but we must not inquire in the classical way for the meaning of these functions by associating them with a property or part of a model, only familiarity with the new concepts will help us.

CHAPTER VI

THE FIRST ORDER EQUATIONS OF THE QUANTUM THEORY

IN the electromagnetic theory of light we have one of the most familiar examples of a wave equation. This may be written in the form

$$\frac{\partial^2\psi}{\partial x^2}+\frac{\partial^2\psi}{\partial y^2}+\frac{\partial^2\psi}{\partial z^2}-\frac{1}{c^2}\frac{\partial^2\psi}{\partial t^2}=0 \quad . \quad . \quad (6\cdot1)$$

In this case ψ denotes any of the components of electric or magnetic intensity.

The interpretation placed upon this equation in the classical electromagnetic theory is that the electric and magnetic intensities are propagated through space with the speed c.

We can also regard (6·1) as the wave equation of a photon. In the wave theory of matter we have made the wave equation the point of departure, but in the electromagnetic theory (6·1) is derived from Maxwell's equations for empty space.

There is a set of six equations of the first order in differential coefficients. We will write them in the vector form as follows :

$$\left.\begin{aligned}\operatorname{curl}\mathbf{E}&=-\frac{1}{c}\frac{\partial\mathbf{H}}{\partial t}\\ \operatorname{curl}\mathbf{H}&=\frac{1}{c}\frac{\partial\mathbf{E}}{\partial t}\end{aligned}\right\} \quad . \quad . \quad (6\cdot2)$$

and with these are associated the two equations :

$$\left.\begin{aligned}\operatorname{div}\mathbf{E}&=0\\ \operatorname{div}\mathbf{H}&=0\end{aligned}\right\} \quad . \quad . \quad . \quad (6\cdot3)$$

Equation (6·1) can be derived from (6·2) and (6·3) in the usual way.

If we imagine that (6·1) was discovered first and that Maxwell's equations were derived as first order equations from which (6·1) is deducible, we have the order of development in the quantum theory.

A similarity between the equations of this theory and the equations of the electromagnetic theory was pointed out early in the history of the quantum theory, but application was made only to the simplest example, viz. that of a particle in the absence of a field of force. It remained to Dirac to find the first order equations by a process which has nothing to do with the analogy with Maxwell's equations.

This analogy has been worked out in detail by several writers, and is exact if the description is made by means of five co-ordinates, i.e. by adopting a five-dimensional system.

If we keep to the four-dimensional scheme of space and time the method adopted is symbolical, and depends upon the introduction of arbitrary operators. This method, which has been followed by Darwin and Frenkel, leads quite simply to Dirac's first order equations. These equations can be regarded as of the same importance in quantum mechanics as those of Maxwell in the electromagnetic theory. Their discovery begins a new stage in the development of the new theory, and in many ways changes the outlook upon it.

In order to understand more easily the procedure in the quantum theory, we will begin with the familiar equations of the electromagnetic theory. When written in terms of the vector components (6·2) and (6·3) give the following eight equations:

$$\frac{\partial \mathrm{E}_z}{\partial y} - \frac{\partial \mathrm{E}_y}{\partial z} + \frac{1}{c}\frac{\partial \mathrm{H}_x}{\partial t} = 0$$

$$\frac{\partial \mathrm{E}_x}{\partial z} - \frac{\partial \mathrm{E}_z}{\partial x} + \frac{1}{c}\frac{\partial \mathrm{H}_y}{\partial t} = 0$$

$$\left.\begin{aligned}
\frac{\partial E_y}{\partial x} - \frac{\partial E_x}{\partial y} + \frac{1}{c}\frac{\partial H_z}{\partial t} &= 0 \\
\frac{\partial H_z}{\partial y} - \frac{\partial H_y}{\partial z} - \frac{1}{c}\frac{\partial E_x}{\partial t} &= 0 \\
\frac{\partial H_x}{\partial z} - \frac{\partial H_z}{\partial x} - \frac{1}{c}\frac{\partial E_y}{\partial t} &= 0 \\
\frac{\partial H_y}{\partial x} - \frac{\partial H_x}{\partial y} - \frac{1}{c}\frac{\partial E_z}{\partial t} &= 0 \\
\frac{\partial E_x}{\partial x} + \frac{\partial E_y}{\partial y} + \frac{\partial E_z}{\partial z} &= 0 \\
\frac{\partial H_x}{\partial x} + \frac{\partial H_y}{\partial y} + \frac{\partial H_z}{\partial z} &= 0
\end{aligned}\right\} \quad . \quad (6.4)$$

Let us write

$$\begin{aligned}
H_x + iH_y = \psi_1, \quad H_z &= \psi_2 \\
-iE_x + E_y = \psi_3, \quad -iE_z &= \psi_4.
\end{aligned}$$

By adding the first of this set of equations to the second, after multiplying the latter by $i = \sqrt{-1}$, we obtain

$$\left(\frac{\partial}{\partial x} + i\frac{\partial}{\partial y}\right)\psi_4 - \frac{\partial}{\partial z}\psi_3 + \frac{1}{c}\frac{\partial}{\partial t}\psi_1 = 0 \quad . \quad (6{\cdot}5)$$

In a similar manner, by combining the above equations in pairs we can obtain three more like (6·5) and replace the eight equations of Maxwell by the following four equations :

$$\left.\begin{aligned}
\left(\frac{\partial}{\partial x} - i\frac{\partial}{\partial y}\right)\psi_3 + \frac{\partial}{\partial z}\psi_4 + \frac{1}{c}\frac{\partial}{\partial t}\psi_2 &= 0 \\
\left(\frac{\partial}{\partial x} + i\frac{\partial}{\partial y}\right)\psi_4 - \frac{\partial}{\partial z}\psi_3 + \frac{1}{c}\frac{\partial}{\partial t}\psi_1 &= 0 \\
\left(\frac{\partial}{\partial x} - i\frac{\partial}{\partial y}\right)\psi_1 + \frac{\partial}{\partial z}\psi_2 + \frac{1}{c}\frac{\partial}{\partial t}\psi_4 &= 0 \\
\left(\frac{\partial}{\partial x} + i\frac{\partial}{\partial y}\right)\psi_2 - \frac{\partial}{\partial z}\psi_1 + \frac{1}{c}\frac{\partial}{\partial t}\psi_3 &= 0
\end{aligned}\right\} \quad . \quad (6{\cdot}6)$$

This group of four equations containing the four ψ-functions may be described as the electromagnetic equations in Dirac's form, since his first order equations of the quantum theory are of this type. In order to obtain them we have to introduce a symbolic method consisting in the replacement of the operators $\frac{\partial}{\partial x}, \frac{\partial}{\partial y}, \frac{\partial}{\partial z}, \frac{\partial}{\partial t}$, by more complicated operators containing the components of the electromagnetic potentials, the vector potential with components (A_x, A_y, A_z), and the scalar potential ϕ. These are introduced into the electromagnetic theory through the equations [1]

$$\left.\begin{aligned} \mathbf{E} &= -\operatorname{grad}\phi - \frac{1}{c}\frac{\partial \mathbf{A}}{\partial t} \\ \mathbf{H} &= \operatorname{curl}\mathbf{A} \end{aligned}\right\} \quad . \quad . \quad (6\cdot7)$$

We saw on page 45 that Schrödinger's equation could be obtained by the substitution of the components p_1, p_2, p_3, by $\frac{h}{2\pi i}\frac{\partial}{\partial x}, \frac{h}{2\pi i}\frac{\partial}{\partial y}, \frac{h}{2\pi i}\frac{\partial}{\partial z}$.

In the presence of the electromagnetic field we have to replace $\frac{h}{2\pi i}\frac{\partial}{\partial x}$ by $u_x = \frac{h}{2\pi i}\frac{\partial}{\partial x} - \frac{e}{c}A_x$, and there are similar substitutions for the y and z components of the operator.

In addition, $\frac{h}{2\pi i c}\frac{\partial}{\partial t}$ must be replaced by

$$u_t = \frac{1}{c}\left(\frac{h}{2\pi i}\frac{\partial}{\partial t} + e\phi\right).$$

We shall see that there is a further slight modification in the substitutions which we shall omit for the present.

[1] See for example, *The Electron Theory of Matter*, O. W. Richardson, 2nd ed., p. 194.

These substitutions enable us to pass at once from Schrödinger's equation for a particle for the case when it is subject to no forces to the case of a particle of mass m carrying a charge e in an electromagnetic field. In the case of the free particle we have (p. 35)

$$W = mc^2 \quad . \quad . \quad . \quad (6{\cdot}8)$$

and the wave equation (2·20) is

$$\frac{\partial^2\psi}{\partial x^2} + \frac{\partial^2\psi}{\partial y^2} + \frac{\partial^2\psi}{\partial z^2} = \frac{G^2}{W^2}\frac{\partial^2\psi}{\partial t^2} \quad . \quad . \quad (6{\cdot}9)$$

From (2·21)

$$\frac{\partial^2\psi}{\partial t^2} = -\frac{4\pi^2}{h^2}W^2\psi \quad . \quad . \quad (6{\cdot}10)$$

Hence

$$\frac{\partial^2\psi}{\partial x^2} + \frac{\partial^2\psi}{\partial y^2} + \frac{\partial^2\psi}{\partial z^2} - \frac{1}{c^2}\frac{\partial^2\psi}{\partial t^2} = \frac{4\pi^2}{h^2}m_0{}^2c^2\psi \quad (6{\cdot}11)$$

Thus by the change of operators we find for the equation of the charged particle in an electromagnetic field

$$(u_x{}^2 + u_y{}^2 + u_z{}^2 - u_t{}^2 + m_0{}^2c^2)\psi = 0 \quad . \quad (6{\cdot}12)$$

If the double operation $u_x{}^2\psi$ be written out in detail, and if we remember that the scalar and vector potentials satisfy the relation

$$\frac{\partial A_x}{\partial x} + \frac{\partial A_y}{\partial y} + \frac{\partial A_z}{\partial z} + \frac{1}{c}\frac{\partial\phi}{\partial t} = 0, \quad . \quad (6{\cdot}13)$$

(6·12) may be written

$$\frac{\partial^2\psi}{\partial x^2} + \frac{\partial^2\psi}{\partial y^2} + \frac{\partial^2\psi}{\partial z^2} - \frac{1}{c^2}\frac{\partial^2\psi}{\partial t^2} - \frac{4\pi ie}{hc}\left(A_x\frac{\partial\psi}{\partial x} + A_y\frac{\partial\psi}{\partial y} + A_z\frac{\partial\psi}{\partial z} + \frac{\phi}{c}\frac{\partial\phi}{\partial t}\right) - \frac{4\pi^2e^2}{h^2c^2}\left(A_x{}^2 + A_y{}^2 + A_z{}^2 - \phi^2 + \frac{m_0{}^2c^4}{e^2}\right)\psi = 0, \quad (6{\cdot}14)$$

$$\left(-\frac{h^2}{\quad}\right).$$

This equation is applicable to a charge, e.g. an electron, without 'spin.' This limitation makes it necessary to seek for a more general equation to take account of the phenomena which hitherto have been explained by supposing an electron to rotate about its axis, i.e. to spin.

Dirac's equations offer a solution to the problem raised.

Maxwell's equations are linear in the operators, but in (6·4) it is to be noted that they are rather restricted in type for in each equation one operator is missing. It will be as well to begin without this restriction. We shall introduce a vector, **M**, with components (M_x, M_y, M_z), together with a vector, **N**, with components (N_x, N_y, N_z), and two scalars M_0 and N_0. Then by substituting $\frac{h}{2\pi i}\frac{\partial}{\partial x}$ by u_x, etc., we obtain a set of equations corresponding to, and suggested by (6·4):

$$\begin{aligned}
u_yM_z - u_zM_y - u_tN_x &= u_xM_0 \quad (a)\\
u_zM_x - u_xM_z - u_tN_y &= u_yM_0 \quad (b)\\
u_xM_y - u_yM_x - u_tN_z &= u_zM_0 \quad (c)\\
u_yN_z - u_zN_y + u_tM_x &= u_xN_0 \quad (d)\\
u_zN_x - u_xN_z + u_tM_y &= u_yN_0 \quad (e)\\
u_xN_y - u_yN_x + u_tM_z &= u_zN_0 \quad (f)\\
u_xM_x + u_yM_y + u_zM_z &= u_tN_0 \quad (g)\\
u_xN_x + u_yN_y + u_zN_z &= -u_tM_0 \quad (h)
\end{aligned} \qquad (6\cdot15)$$

The intensities in Maxwell's equations are replaced by the new vectors **M** and **N**.

From this group of equations, by analogy with Maxwell's equations, we may expect to derive a second order equation similar to (6·14). It is possible to do this, and it may be shown that the equation derived is very nearly identical with this equation, but differs from it in two important points. In the first place, additional terms occur which suggest the presence of a magnetic and an electric dipole, the former being required by the 'spin' theory. In the second place, the term containing m_0^2 is missing. This term is essential in the theory, for the

equation derived must reduce to (6·11) in the absence of an electromagnetic field, and it cannot do this unless the mass term is present.

It is possible to introduce this term by replacing u_t by another operator.

In (*a*), (*b*), (*c*) and (*g*) u_t is replaced by $(u_t - m_0c)$, and in the remaining equations it is replaced by $(u_t + m_0c)$, where m_0 is the rest mass of the charged particle.

This process is not so arbitrary as may appear at first sight. We are guided by the analogy with Maxwell's equations and by the fact that the second order equation must become (6·11) when there is no field. This gives the clue to the modification necessary in the group (6·15), and there is no alternative to the process just described. The set of equations we have obtained can be combined by a process similar to that by which (6·4) were combined.

We make the substitutions

$$\left.\begin{aligned} \psi_1 &= \mathrm{M}_x + i\mathrm{M}_y, & \psi_2 &= \mathrm{M}_z - i\mathrm{M}_0 \\ \psi_3 &= -i\mathrm{N}_x + \mathrm{N}y, & \psi_4 &= -i\mathrm{N}z - \mathrm{N}_0 \end{aligned}\right\} \quad (6\cdot16)$$

If $u_t' = u_t - m_0c$ and $u_t'' = u_t + m_0c$ we obtain

$$\left.\begin{aligned} (u_x - iu_y)\psi_3 + u_z\psi_4 + u_t''\psi_2 &= 0 \\ (u_x + iu_y)\psi_4 - u_z\psi_3 + u_t''\psi_1 &= 0 \\ (u_x - iu_y)\psi_1 + u_z\psi_2 + u_t'\psi_4 &= 0 \\ (u_x + iu_y)\psi_2 - u_z\psi_1 + u_t'\psi_3 &= 0 \end{aligned}\right\} . \quad (6\cdot17)$$

These are Dirac's equations which, however, he obtained by a method which is quite different from that introduced here.

Four ψ-functions replace the single function of Schrödinger's equation.

The Second Order Equation

We now proceed to find the second order equation which corresponds to (6·1). There is an equation for each component of **M** and of **N** as well as for M_0 and N_0.

We obtain the equation in M_0.

Before passing to this, let us examine the operation

$$(u_y u_z - u_z u_y)\mathrm{M}_0$$

$$u_y u_z \mathrm{M}_0 = \left(\frac{h}{2\pi i}\frac{\partial}{\partial y} - \frac{e}{c}\mathrm{A}_y\right)\left(\frac{h}{2\pi i}\frac{\partial \mathrm{M}_0}{\partial z} - \frac{e}{c}\mathrm{A}_z\mathrm{M}_0\right)$$

$$= \left(\frac{h}{2\pi i}\right)^2 \frac{\partial^2 \mathrm{M}_0}{\partial y\,\partial z} - \frac{he}{2\pi i c}\frac{\partial}{\partial y}(\mathrm{A}_z\mathrm{M}_0) - \frac{he}{2\pi i c}\mathrm{A}_y\frac{\partial \mathrm{M}_0}{\partial z} + \left(\frac{e}{c}\right)^2 \mathrm{A}_y\mathrm{A}_z\mathrm{M}_0.$$

Hence

$$(u_y u_z - u_z u_y)\mathrm{M}_0 = \frac{he}{2\pi i c}\left(\frac{\partial \mathrm{A}_y}{\partial z} - \frac{\partial \mathrm{A}_z}{\partial y}\right)\mathrm{M}_0 = -\frac{he}{2\pi i c}\mathrm{H}_x\mathrm{M}_0$$

by (6·7). (6·18)

Similar relations hold for

$$(u_z u_x - u_x u_z) \text{ and } (u_x u_y - u_y u_x).$$

In the same way we find

$$(u_x u_t - u_t u_x)\mathrm{M}_0 = -\frac{he}{2\pi i c}\mathrm{E}_x\mathrm{M}_0, \quad . \quad (6·19)$$

with similar relations for the y and z components.

In deriving the second order equation from (6·15) we must remember the change made in u_t. The first equation (a) of the group is now

$$u_y\mathrm{M}_x - u_z\mathrm{M}_y - u_x\mathrm{N}_t + m_0 c\mathrm{N}_x = u_x\mathrm{M}_0$$

and the last (h) is

$$u_x\mathrm{N}_x + u_y\mathrm{N}_y + u_z\mathrm{N}_z = -u_t\mathrm{M}_0 - m_0 c\mathrm{M}_0.$$

Operate upon (a), (b) and (c) by u_x, u_y and u_z respectively and upon (h) by $(u_t - m_0 c)$ and add together the four equations obtained.

The second order equation which results is

$$(u_x^2 + u_y^2 + u_z^2 - u_t^2 + m_0^2 c^2)\mathrm{M}_0$$
$$= (u_y u_z - u_z u_y)\mathrm{M}_x + (u_z u_x - u_x u_z)\mathrm{M}_y + (u_x u_y - u_y u_x)\mathrm{M}_z$$
$$- (u_x u_t - u_t u_x)\mathrm{N}_x - (u_y u_t - u_t u_y)\mathrm{N}_y - (u_z u_t - u_t u_z)\mathrm{N}_z$$
$$= -\frac{he}{2\pi i c}(\mathrm{H}_x\mathrm{M}_x + \mathrm{H}_y\mathrm{M}_y + \mathrm{H}_z\mathrm{M}_z - \mathrm{E}_x\mathrm{N}_x - \mathrm{E}_y\mathrm{N}_y - \mathrm{E}_z\mathrm{N}_z) \quad . \quad . \quad (6·20)$$

The equation in N_0 is similar to (6·20) and becomes identical with it if we write $N_0 = \pm iM_0$. In the same way the equations in M_x and N_x become identical if $N_x = \pm iM_x$ and similarly for the y and z components. We can write for the vectors **M** and **N**

$$\mathbf{N} = \pm i\mathbf{M} \quad . \quad . \quad . \quad (6\cdot21)$$

The existence of a real magnetic moment is suggested by experiment, and we are led to suppose that the electric moment must be imaginary, since if **M** is real **N** is imaginary by (6·21).

The terms on the right-hand side of (6·20) are extra terms which have entered quite naturally as a result of the derivation of the second order equation from (6·15). Their form is suggestive, for we know that if a magnet of moment **M**, the components being M_x, M_y, M_z, is situated in a field of intensity **H**, with components H_x, H_y, H_z, the energy of the magnet is

$$(H_xM_x + H_yM_y + H_zM_z).$$

Similarly, an electric doublet $-$ **N**, situated in a field of intensity **E**, has energy $-(E_xN_x + E_yN_y + E_zN_z)$.

If we refer to Schrödinger's equation (2·25), we see that the energy terms have a factor $\frac{8\pi^2m_0}{h^2}\psi$. Let us then introduce a magnetic moment **m**, with components m_x, m_y, m_z, and an electric moment **n**, with components n_x, n_y, n_z, so that the extra terms when taken over to the left-hand side are reduced to

$$\frac{8\pi^2m_0}{h^2}(H_xm_x+H_ym_y+H_zm_z+E_xn_x+E_yn_y+E_zn_z)M_0.$$

The left-hand side of (6.20) is equal to $\frac{h^2}{4\pi^2}$ times the left-hand side of (6·14). For the sake of shortness let the left-hand side of (6·14) be denoted by $D\psi$. Then (6·20) can be written

$$DM_0 + \frac{2\pi i}{h}\frac{e}{c}(H_xM_x + \ldots - E_xN_x + \ldots) = 0,$$

$$\text{or } \mathrm{DM}_0 + \frac{8\pi^2 m_0}{h^2}(\mathrm{H}_x m_x + \ldots + \mathrm{E}_x n_x + \ldots)\mathrm{M}_0 = 0, \tag{6·22}$$

where $m_x = \dfrac{he}{4\pi m_0 c}\, i\dfrac{\mathrm{M}_x}{\mathrm{M}_0}, \quad n_x = -\dfrac{he}{4\pi m_0 c}\, i\dfrac{\mathrm{N}_0}{\mathrm{M}_x}.$

The second order equation which we have now obtained thus suggests that in addition to the energy associated with a moving charge we must add that of a magnet of moment, **m**, and that of an electric doublet of moment, **n**.

The spinning electron introduced already into the atomic model generates a magnetic moment by its rotation. Hitherto we have not introduced an electric doublet into the atomic model, and we have noticed that if the magnetic moment is real the electric moment is imaginary. We may perhaps regard this imaginary electric moment as possessing no physical significance.

We must beware of requiring an exact model to correspond with the theory. The view now is that it is impossible to obtain pictures of the fundamental things in physics. Nevertheless, it is surprising how closely the theory we have been considering corresponds to the familiar atomic model.

It is possible to deduce from (6·22) that the absolute value of **m** is $\dfrac{he}{4\pi m_0 c}$, and that it can have two directions, one along the magnetic field and the other opposite to it.

This is exactly the value obtained by Goudsmit and Uhlenbeck in their theory of the spinning electron.

A number of difficulties had arisen in the theory of Bohr and Sommerfeld in connection with the explanation of the anomalous Zeeman effect.

Goudsmit and Uhlenbeck succeeded, on the whole, in removing these by supposing that the electron had a spin and their calculated results when compared with spectroscopic data led to the above properties of **m**.

Thus the problem of the atom must be approached,

not by applying the original Schrödinger equation (2·22) or the approximate form of it (2·25), but by beginning directly with the equations we have just developed. We have seen that (2·25) gives the correct energy levels for the hydrogen atom (p. 61), and (2·22), being the more general form, might be expected to account for the fine structure of the hydrogen lines by substituting the relativity value of G.

In the old quantum theory the two cases were treated by applying the quantum laws to the atom model, firstly, assuming the mass to be constant, and secondly, taking into account the relativity variation of mass with velocity.

In the first case the simple series of energy values was obtained, and in this respect the old and new theories are in agreement.

In the second case, Sommerfeld obtained fine structure formulæ which were in agreement with experiment in the case of hydrogen and ionized helium. There was good agreement also in the case of X-ray spectra, but an error occurred in the number of lines; for example, in the L-group Sommerfeld's theory suggested that the number of lines was two, while the actual number is three, the numbers being deduced by the application of his selection principle.

The work of Millikan and Bowen on 'stripped atoms' led them to conclude that the splitting of the energy levels in the X-ray spectra and in atoms other than those of hydrogen and helium was of magnetic origin.

The situation then appeared to be that in the case of hydrogen and helium the splitting of the levels was relativistic and in the other atoms magnetic in origin, and it was by chance that the formulæ for the two different cases were similar. When the second case is treated by means of Schrödinger's equation the result obtained is incorrect. The reason is that no account is taken of electron 'spin.'

When we begin with the new set of equations (6·17) the correct result is obtained. At present this is the

most satisfactory method, both for practical and theoretical reasons. No special assumptions are made, the basis of the derivation of the equations being that a set of first order equations similar to those of Maxwell govern the quantum phenomena.

The actual working out of the formulæ is long, and we shall have to refer to larger works for the details.

The Universe as a Five-Dimensional Continuum

We shall bring our considerations to an end with a short account of another point of view to which we are led by the recent developments in atomic physics. The goal here is a unitary theory of physics, that is the discovery of a method of description which combines into a united whole all the diverse phenomena of inanimate nature. The development of classical theories led in this direction. We see the attempt to discover a unifying principle in mechanics, thermodynamics, and in all branches of physics. The electromagnetic theory represents the culmination of the classical striving for the discovery of a single natural principle. The theory of relativity, in which we include the metrical theory of Weyl and Eddington, seems to bring the goal within reach, but always quantum phenomena stand aside. This applies to the later developments of all four-dimensional theories of the universe.

It appears that by adopting a five-dimensional system of reference one can approach this ideal even to the extent of including quantum phenomana. The goal is not yet attained, but we appear to have advanced so far on the way that our steps can hardly be all in error.

In Einstein's theory of relativity, gravitation became recognized as a geometrical property of space. The phenomena of gravitation proceeded from the assumption that natural space is four-dimensional and is characterized by the system of measurement

$$ds^2 = g_{mn}dx^m dx^n \quad . \quad . \quad (6{\cdot}23)$$

We need not explain the notation since it is now so ıidely known. The result is to banish the old idea of ravitational force, which becomes no more than an ıdication that the universe is not 'flat' and therefore ot to be described by means of the geometry of Euclid. f we persist in using this geometry we obtain an incorrect iew of the phenomena of gravitation and have to intro- uce the conception of force.

From this scheme electromagnetic phenomena appear o stand apart. Electric and magnetic forces still remain ıninterpreted as geometrical properties.

As has been mentioned above, the equation (6·23) is haracteristic of Riemannian space, which is the basis of he theory of relativity. In Weyl's theory we no longer uppose that the system of measurement at every point s given by (6·23). At each point it is not the absolute ralues of the g's which have physical significance, but ɔnly their ratios, and with this assumption Weyl sup- ɔoses that a length l of a vector at any point, undergoes ı change of length dl in a parallel displacement to a ıeighbouring point, given by

$$\frac{dl}{l} = -\phi_i dx^i \quad . \quad . \quad . \quad (6\cdot24)$$

$$l_o = le^{\int \phi_i dx^i} \quad . \quad . \quad . \quad (6\cdot25)$$

In general the integral depends on the path of integration, so that if we begin with a length l_0 at A we obtain a different length l at B according to the path taken from A to B. Weyl identifies the function ϕ_i with a component of the electromagnetic vector potential (see 6·7).

Thus Weyl sees in electromagnetic phenomena the expression of this metrical property of space.

The length can, of course, denote a duration in time, a standard period, such as that occurring in atomic processes, and which we recognise in the sharpness of spectral lines. This period appears to be independent of the previous history, and this fact seems to

indicate that natural processes do not follow Weyl's law. It appears that Weyl's suggestion should properly be applied to a five-dimensional continuum, and that the function ϕ_i should be identified, not with the electromagnetic potential, but with the generalized momentum $\left(p_i + \frac{e}{c}\phi_i\right)$, where p_i, as before, denotes the mechanical momentum and e the fundamental electric charge.

The theory of wave mechanics indicates that Weyl's law should be replaced by

$$\frac{d\mathrm{L}}{\mathrm{L}} = \frac{2\pi i}{h} \Pi_\mu dx^\mu \qquad . \qquad . \quad (6{\cdot}26)$$

where L is a five-dimensional length and Π_μ is a typical component of generalised momentum with $\Pi_5 =$ a constant, which can be identified with the mass of the particle, of which the momentum is p_i, and which carries a charge e.

We cannot go into this in detail, and are here merely mentioning the results to which the theory seems to force us. In such a continuum, when a vector function varies from point to point, we have always to remember that there is a change, due to the displacement, which is a consequence of the varying metrical properties. This natural change modifies the expressions such as $\frac{\partial}{\partial x}$ and $\frac{\partial}{\partial t}$ with which we are familiar in ordinary space and time variation, and it alters also the variations employed in the space-time of relativity.

Thus an expression like div A is rather more complex than it is in the usual sense in which the operator 'div' is employed. In the five-dimensional continuum of Kaluza and Klein we write instead of (6·23)

$$d\sigma^2 = \gamma_{\mu\nu} dx^\mu dx^\nu, \qquad . \qquad . \qquad . \quad (6{\cdot}27)$$

where μ and ν have the values 1 to 5.

The values of the γ's depend upon the g's of Einstein and the functions ϕ_i, the electromagnetic potentials, e.g.

$$\gamma_{ik} = g_{ik} + \gamma_{55}\alpha^2\phi_i\phi_k, \quad \gamma_{5i} = \alpha\gamma_{55}\,\phi_i, (i = 1, 2, 3, 4) \quad (6{\cdot}28)$$

where γ_{55} and α are constants.

The shortest paths in this space, or the geodesics, represent the paths of charged particles in a gravitational and an electromagnetic field, just as in Einstein's theory the geodesics are the paths of particles in a gravitational field.

Thus in this system the electromagnetic phenomena are placed upon the same geometrical footing as the gravitational.

The momentum, Π_μ, of which the significance was pointed out by Professor W. Wilson,[1] has a very simple significance in the continuum. Just as p_i is defined in four dimensions by $p_i = m_0 g_{ik}\frac{dx^k}{ds}$, so Π_μ is defined here by $\mathrm{I}\gamma_{\mu\nu}\frac{dx^\nu}{d\sigma}$, where I corresponds in five dimensions to the mass m_0 in four.

It has been shown by J. W. Fisher [2] that if we write $\gamma_{55} = -1$ and $\alpha = \frac{e}{m_0 c}$, (6·27) becomes zero and thus the track of a particle in a gravitational and an electromagnetic field is a null geodesic. At the same time the value of I becomes zero, but the ratio $\frac{d\sigma}{\mathrm{I}}$ remains definite, and equal to $\frac{ds}{m_0}$, where ds is the four dimensional element of (6·23). This makes the particle in five dimensions exactly analogous to a photon in four dimensions, for a ray of light, which is the track of a photon, is a null geodesic in the four-dimensional continuum and in wave mechanics the mass of a photon is zero.

[1] *Roy. Soc. Proc.*, A, vol. 102, 1922, p. 478.
[2] *Ibid.*, vol. 123, p. 490.

We are tempted at once to write down the wave equation in five dimensions which corresponds to the light wave in four. It is very striking that this five-dimensional wave equation is Schrödinger's relativity equation (6·12) or (6·14). It is thus the equation for a non-rotating electron.

But we must not hurry too rapidly to this derivation of a wave equation by analogy with Riemannian space. We must remind ourselves of the variations that occur in the metric, which we neglect in the analogous derivation just mentioned.

We must proceed by writing down the equations analogous to those of Maxwell for empty space (6·2) and (6·3). If we do this, having regard to the metrical variation, we obtain (6·15) in the final form, i.e. with the term in $m_0 c$. This comes about without any introduction of symbolical methods.

Thus the fundamental quantum equations may be regarded as indicating the need to adopt a particular metric in a five-dimensional continuum, the special metrical function being Π_μ, the generalised momentum.

These considerations may well indicate that the true basis of the wave-like character of matter is the five-dimensional null geodesic.

Another point of interest is that out of the continuum we select only null geodesics as having physical significance.

This, together with the fact that we persistently describe nature in the four-dimensional continuum of space and time, may be the basis of our discovery of discontinuity in nature.

It is remarkable in this theory that e appears as a single constant. It does not possess a positive and a negative value. If a particle is travelling so that it has a positive fifth component of momentum it appears in the equation as a positive charge, if the component is negative it appears as a negative charge. Thus the difference in the sign of the fifth component of momentum is responsible for our perception of two different kinds

of electrical charge. By substitution of the value of Π_μ in (6·26), it follows that along a null geodesic, the change dL is zero. But when we describe our quantities in terms of four dimensions, it does not follow that the four-dimensional dl vanishes. In fact, dL is made up of two parts, viz. dl and a component depending upon the fifth co-ordinate, and $\frac{dl}{l}$ is a periodic function of the fifth co-ordinate, the period being $\frac{h}{m_0}$. Now it is a remarkable fact that the value of $\frac{dl}{l}$ appears to be different for the proton and electron. In the former case the periodicity in the fifth component is $\frac{h}{M_0}$, where M_0 is the mass of the proton, while in the latter it is $\frac{h}{m_0}$, where m_0 is the mass of the electron.

Thus the change in length per unit length is much more rapid along the proton path than along the electron path, and the difference geometrically is one of direction along the fifth axis. This seems to suggest that there is a difference in natural metric in the two directions along this axis, 'equal' lengths have 'different' values in these two directions. We could interpret this by saying that the same physical effect is experienced along different lengths in the two directions. Thus a force applied to a proton causes it to be accelerated to a certain extent, measured in the special units for the proton, while the same force applied to an electron produces the same acceleration, but measured in the units appropriate to the electron. It is thus our persistence in adopting the same units for both which is responsible for the apparent difference in mass in these two bodies.

This scheme of geometry and metrics offers the possibility of the attainment of a remarkable unifying principle in physics and invites further exploration.

REFERENCES

G. Birtwistle. The New Quantum Mechanics. (Camb. Univ. Press, 1928.)

N. Bohr. Das Quantenpostulat und die neuere Entwicklung der Atomistik. (*Naturwissenschaften*, Art. 453.)

N. Bohr. The Quantum Postulate, and the Recent Development of Atomic Theory. (*Nature*, Vol. 121, Ap. 1928, p. 589.)

L. de Broglie. Sur le parallelisme entre la dynamique du point materiel et l'optique géométrique. (*J. de Physique*, 6e Série, t. vii, 1926, p. 1.)

L. de Broglie. Introduction to Wave Mechanics. (Methuen.)

F. Brouwer. Wellenmechanische Eigenwertprobleme. (*Ann. d. Physik*, 84, p. 915.)

P. A. M. Dirac. Roy. Soc. Proc., A, 117, p. 610. Quantum Mechanics. (Clarendon Press.)

de Donder et van den Dungen. La quantification déduite de la gravifique einsteinienne. (*C.R. Acad. Sci.*, t. 183, 1926, p. 22.)

L. Flamm. Die Grundlage der Wellenmechanik. (*Phys. Zeitscher.*, 27 Jahrg., p. 600.)

J. Frenkel. Einführung in die Wellenmechanik. (Berlin, Julius Springer.)

W. Heisenberg. Uber den anschaulichen Inhalt U.S.W. (*Z. f. Physik*, 43, 1927, p. 172.)

Th. Kaluza. Zum Unitätsproblem der Physik. (*Berliner Sitzungsberichte*, 1921, p. 966.)

O. Klein. Quantentheorie und fünf dimensionale Relativität. (*Z. f. Physik*, 37, 1926, p. 895.)

O. W. Richardson. On the Present State of Atomic Physics. (*Proc. Phys. Soc.*, 39, p. 171.)

O. W. Richardson and H. T. Flint. On a Minimum Proper Time and Applications. (*Roy. Soc. Proc.*, A, 117, p. 637.)

L. Rosenfeld. L'Univers à cinq dimensions et la Mécanique Ondulatoire. (*Acad. Roy. de Belgique Sci.*, 5e ser., t. xiii, 6, p. 1.)

E. Schrödinger. Quantisierung als Eigenwert problem. (*Ann. d. Physik*, 79, 1926, p. 489.) Über das Verhältnis des Heisenberg - Born - Jordansche Quantenmechanik zu der meinen. (*Ann. d. Physik*, 79, 1926, p. 734.) Der stetige Übergang von der Mikro- zur Makromechanik. (*Naturwissenchaften*, 14 Jahrg., p. 664.)

G. P. Thomson. Experiments on the Diffraction of Cathode Rays. (*Roy. Soc. Proc.*, A, 117, p. 600.)

W. Wilson. Relativity and Wave Mechanics. (*Roy. Soc. Proc.*, A, 118, p. 441.)

B. L. Worsnop. Note on the Transmission of Cathode Rays through Thin Films. (*Proc. Phys. Soc.*, June, 1928.)

INDEX

PRINTED IN GREAT BRITAIN AT THE UNIVERSITY PRESS, ABERDEEN